Aus dem Institut für Schlaganfall- und Demenzforschung
Klinikum der Universität München, Großhadern
Direktor: Prof. Dr. Martin Dichgans

Automated Detection of Early-Stage Alzheimer's Disease

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Jinyi Ren
aus Shandong, China
2020

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. Michael Ewers
Mitberichterstatter:	Prof. Dr. Thomas Pfluger
	PD Dr. Markus Pfirrmann
	Prof. Dr. Andreas Straube
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	24.03.2020

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

„Automated Detection of Early-Stage Alzheimer's Disease"

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 27.04.2020	Jinyi Ren
Ort, Datum	Unterschrift

Table of contents

Abbreviations

AD – Alzheimer’s Disease

SAD – Sporadic Alzheimer’s disease

ADAD – Autosomal Dominant Alzheimer’s disease

CSF – Cerebrospinal Fluid

Aβ – beta-amyloid plaques

pTau – hyperphosphorylated tau

NFT – neurofibrillary tangles

MRI – Magnetic Resonance Imaging

FDG-PET – Fluordesoxyglucose Positron Emission Tomography

MCI – Mild Cognitive Impairment

HC – Healthy Control

Cov - Converter

ADNI – Alzheimer's Disease Neuroimaging Initiative

DIAN – Dominantly Inherited Alzheimer’s Network

EYO – estimated years to symptom onset

PSEN1 – presenilin 1

PSEN2 – presenilin 2

APP – amyloid precursor protein

AUC – Area Under Curve;

SEN – Sensivity;

SPE – Specificity

SUVR – standardized uptake value ratio (SUVR)

ROI – region of interest

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disease ranging from an asymptomatic preclinical phase to dementia. A small subset (less than 5%) of AD patients, starting at an early age (from 30 to 65 years old), was found mainly caused by the genetic mutation which is autosomal dominant AD (ADAD). Even though ADAD only accounts for a small proportion of AD, because of similar pathological and clinical features, it is critical to better understand the pathophysiology in the preclinical stage of sporadic AD (SAD). Decades before the dementia onset, the earliest changes have already occurred in the brain even though the cognitive function is still normal. Such a long phase of gradual brain changes offers the opportunity to detect the disease in the preclinical phase. However, due to the subtle degree of the brain changes and the unclear etiology in the early stage of AD, it is a challenging task to establish markers for the prediction of the development of dementia. Moreover, subjects in the early-stage SAD may have additional brain changes that are unrelated to AD pathology. To address this problem, we proposed a machine learning model based on patients with ADAD, i.e. pure genetically caused AD which not accompanied by aging-related co-pathologies due to the early disease onset that can start in the third decade of life.

Previous studies in both ADAD and SAD have shown that markers of pathological brain changes including brain atrophy, cerebral glucose hypometabolism, as well as amyloid-beta (Aβ) and pathologic tau protein are altered in AD (Clifford R. Jack *et al.*, 2010; Dickerson and Wolk, 2013). Recent guidelines for early diagnosis of AD recommended using a combination of neuroimaging, CSF biomarkers and neuropsychological tests for aiding the diagnosis of AD (Sperling *et al.*, 2011). In the present work, we established a classification model for the early detection of AD in the predementia phase using machine learning. First, we employed the Naïve Bayes algorithm for classification and information gain method for

feature selection. Features were derived from neuroimaging (MRI, FDG-PET), CSF biomarkers, and neuropsychological tests. First, we trained the classification model based on the ADAD sample and subsequently validated the best prediction model with optimal combinations of multi-modalities in a larger independently recruited sample of subjects with SAD (ADNI dataset).

We found high accuracy for distinguishing HC vs. AD and MCI converters vs. non-converters. Generally, combining multiple modalities, the classification model yielded a better result than the single-modality model. Particularly, the classification for MCI converters versus non-converters in SAD showed the best result when using FDG-PET, neuropsychological and CSF data with AUC of 89.12% (sensitivity = 82.79%, specificity = 82.42%) which is above the clinical relevant accuracy. In summary, the machine learning model established in DIAN achieved predictive accuracy to distinguish the converters versus non-converters in the early stage of SAD. Moreover, the current study provides good interpretability to understand SAD through ADAD.

Zusammenfassung

Die Alzheimer Krankheit (AK) ist eine fortschreitende neurodegenerative Erkrankung, die ausgehend von einer asymptomatischen präklinischen Phase in eine Demenz mündet. Schon Jahrzehnte bevor klinische Symptome sichtbar werden, zeigen sich pathologische Veränderungen im Gehirn der Betroffenen. Diese lange prodromale Phase bietet die Möglichkeit die Krankheit frühzeitig zu erkennen und eine Therapie einzuleiten. Allerdings sind die ersten Hirnveränderungen oft marginal und auch wegen der unklaren Ätiologie ist es im Frühstadium der AK schwierig eine gute Vorhersage darüber zu treffen, welcher Patient in das nächste Krankheitsstadium fortschreiten wird. Zusätzlich leiden gerade ältere Patienten mit einer sporadischen AK, deren Risiko nach der 6. Lebensdekade start ansteigt, oft an altersbedingten Hirnläsionen, die nicht unmittelbar mit Alzheimer in Verbindung stehen. Bei der genetisch verursachten Form der familialen AK (autosomal-dominant vererbte Alzheimer Krankheit, ADAK) treten die ersten Symptome bereits aber der 3. Dekade auf, also weitaus früher als bei der sporadischen Form der AK). Aufgrund des frühen Krankheitsbeginns, sind die AD-spezifischen Gehirnveränderunge bei der ADAK nicht durch alterskorrelierte Co-patholgien wie zB zerebrovaskuläre Erkrankungen konfundiert. Daher kann die ADAK trotz geringer Häufigkeit (< 5% aller AK Fälle) als Modell dienen, um AD spezifische Marker für die frühe Krankheitserkennung zu etablieren.

Vorausgegangene Studien zeigten, dass Marker der Abnahme des Hirnvolumens (Hirnatrophie), zerebraler Glukose-Hypometabolismus, sowie Ablagerungen seniler Plaques aus Beta-Amyloid Peptiden und Neurofibrillenbündel bestehend aus dem Tau-Protein einen bei der AK sowie ADAK abnorm verändert sind (Clifford R. Jack *et al.*, 2010; Dickerson and Wolk, 2013). Aktuelle Richtlinien zur Alzheimer Frühdiagnose empfehlen, dass eine Kombination aus Bildgebung, Liquorbiomarker und neuropsychologischen Untersuchungen zu einer verbesserten Diagnosestellung hergenommen werden (Sperling *et al.*, 2011). In der

vorliegenden Arbeit kam ein naiver Bayes Algorithmus für die Identifikation der wichtigsten Klassifikationsmerkmale (feature selection) zum Einsatz. Zunächst etablierten wir multimodale Vorhersagemodelle basierend auf den Bildgebungsdaten (MRT, FDG-PET), Liquormarkern und neuropsychologischen Tests in ADAK Patienten aus der DIAN-Studie. Danach wurde das Modell mit der besten Vorhersageleistung ausgewählt, um Probanden mit einer sporadischen AK aus der ADNI-Studie zu klassifizieren. Unser Modell konnte mit hoher Genauigkeit gesunde Kontrollprobanden von Alzheimerpatienten unterscheiden und besonders kritisch progrediente Probanden mit leichter kognitiver Störung – also diejenigen, die zum Demenz-Stadium konvertieren - von stabil bleibenden Probanden trennen.

1 Introduction

1.1. Sporadic and autosomal-dominant Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder ranging from an asymptomatic preclinical phase to dementia (Mattson, 2004). The clinical syndrome is characterized by cognitive impairment including loss of memory, problems with language, executive function, and visuospatial abilities and so on (Wilson *et al.*, 2012). Even though the cause of the disease is unknown, pathological features of AD include loss of neurons, accumulation of amyloid plaques, neurofibrillary tangles (Terry *et al.*, 1991; Touchon and Ritchie, 1999).

Globally, based on the statistic of World Alzheimer Report in 2018, there were approximately 50 million people with dementia (Christina, 2018). By 2040, the number of people who have dementia will be doubled (Prince *et al.*, 2016). The estimated cost for the treatment and care of individuals with dementia in 2018 was a trillion dollars worldwide. In the United States, Alzheimer's disease (AD) was ranked as the second most frequent disease leading to fatal health problems based on one Medicare survey (Mayeux *et al.*, 2011; Prince *et al.*, 2015; Niu *et al.*, 2017).

Sporadic Alzheimer's disease (SAD) is an age-related neurodegenerative disorder, where the risk of SAD-related dementia symptoms starts to increase after the age of 65 years. In a small subset (less than 5%) of AD patients, AD is caused by genetic mutations and is called autosomal dominant (ADAD) (Cruts and Van Broeckhoven, 1998). The known genetic mutations causing ADAD occur in genes encoding the amyloid precursor protein (APP) and the genes for the presenilin 1 (PSEN1) and presenilin 2 (PSEN2) proteins, among which mutation in PSEN1 is most common in ADAD (Campion et al., 1999). Even though ADAD only accounts for a small

proportion of AD, the familial early onset-form of AD provides an excellent opportunity to investigate the course of the disease. A longitudinal multicenter study, i.e. the Dominantly Inherited Alzheimer's Network (DIAN), has been started in 2008 to assess biomarker, neuroimaging and cognitive changes in ADAD. Based on the parental onset of symptom, it gives an opportunity to track the brain changes of asymptomatic offspring (mutation carriers) 10-20 years before the estimated onset (Bateman et al., 2012).

The main pathogenesis in both sporadic and autosomal dominant AD include the accumulation of beta-amyloid plaques (Aβ) and neurofibrillary tangles consisting of hyperphosphorylated tau protein which could be measured in cerebrospinal fluid (CSF) or by positron emission tomography (PET) imaging (Klunk *et al.*, 2004; Buerger *et al.*, 2006; Walsh and Selkoe, 2007; Fodero-Tavoletti *et al.*, 2011). As the main component of amyloid plaque, $A\beta_{1-42}$ is the most important pathological peptide that aggregates and forms diffuse plaques, entailing a cascade of brain alterations such as the activation of astrocytes and microglia, increased development of pathologic tau, and ultimately widespread neuronal and synaptic dysfunction (Blennow, de Leon and Zetterberg, 2006). Due to the aggregation of Aβ deposits in plaque and then less diffusion in CSF, $A\beta_{1-42}$ level can be detected in the CSF, where in AD CSF levels are *reduced* presumably due to the deposition of the Aβ peptide in the plaques and thus reduced availability of soluble Aβ in the CSF (Tapiola *et al.*, 2009).

APP, PSEN1 and PSEN2 are associated with the pathogenesis of the amyloid-β. These three genes are involved in the proteolytic cleavage of the APP protein, leading to the beta-amyloid protein, where the APP and PSEN mutations enhance the production of the Aβ protein (Ryan and Rossor, 2010).

Besides Aβ, phosphorylated tau is another important AD pathology which is the main component of neurofibrillary tangles. Tau is a soluble microtubule-associated protein whose main function is to stabilize intracellular microtubules. Phosphorylated tau aggregates into

paired helical filaments (PHFs), and further forms neurofibrillary tangles (NFTs), eventually causing neuronal death (Lee *et al.*, 2005). The distribution of tau can also refer to the Braak staging which was described by Heiko Braak in 1991 (Braak and Braak, 1991; Braak *et al.*, 2006). The stage 1 and 2 marked as lesion firstly developed in the transentorhinal region then progress to the entorhinal region. Stage 3 and 4 marked as the involvement of hippocampus and entorhinal cortex. Stage 5 and 6 marked as tau widely progress to other neocortex regions. Evidence from Iqbal suggested that he pathological tau protein, like prion protein, can spread through neurons in the brain and eventually affect the remote brain area as the disease progresses (Iqbal *et al.*, 2005). Presumably due to the degeneration of neurons, intraneuronal tau and phosphorylated tau are released into the extracellular space, which becomes detectable in the CSF. CSF total tau is a biomarker of neurodegeneration since it is enhanced also in other neurodegenerative diseases without NFT, whereas phosphorylated tau is often regarded as a marker of neurofibrillary tau pathology (Shaw *et al.*, 2009).

To detect AD-related early changes in the brain, neuroimaging techniques including MRI and PET have been widely used in recent years. The pattern of brain atrophy can be examined by structural MRI via gray matter volume change. AD-related brain atrophy occurs in multiple areas including the preferentially the medial temporal lobe and posterior parietal cortex at an early disease stage and involving other brain areas eventually as well. MRI assessed hippocampus volume is a key neuroimaging marker of AD, that was found to be reduced by 32.5% in AD dementia patients compared to healthy subjects (Frisoni *et al.*, 2008). In MCI patients, there is a significant difference in limbic and fronto-temporo-parietal neocortical atrophy compared to the HC group. The general brain atrophy pattern in MCI by structural MRI is an important indicator to predict the progression from MCI to AD (Misra, Fan and Davatzikos, 2009). The best-established functional neuroimaging marker includes [^{18}F]fluorodeoxyglucos PET (FDG-PET) assessed glucose metabolism in temporoparietal brain regions (Furst *et al.*,

2012). In AD patients, brain metabolism is reduced between 30% to 70% compared with cognitively normal elderly people (Silverman *et al.*, 2001).

1.2. Changes in Alzheimer's disease

Alzheimer's disease is slowly progressive and usually can be classified as three stages based on the severity of cognitive impairment: cognitively normal, mild cognitive impairment and dementia. In the very early stage of AD, there is subtle or no impairment shown in cognitive function, but the AD pathophysiology in the brain is already ongoing. Mild cognitive impairment (MCI) is an intermediate state between normal cognition and dementia and it is always considered as an early stage of AD due to up to two-thirds of patients with MCI underlying AD pathology. Even though there is wide heterogeneity, patients of MCI have a higher risk to progress to dementia.

Many years - even decades - before the symptom onset of AD, pathological changes begin to accumulate in the brain. Understanding the order and rate of pathophysiological changes in the progress of AD is important for preclinical diagnosis. Aβ accumulation starts earlier than other pathologies and becomes abnormal up to 20-25 years before the onset of dementia symptoms (Jansen *et al.*, 2015). Following Aβ, Synaptic dysfunction appears in the preclinical stage of AD which can be accessed by [^{18}F]fluorodeoxyglucos PET (FDG-PET) and functional MRI (fMRI). FDG-PET hypometabolism starts to develop in the brain especially in the temporoparietal cortex. Following FDG hypometabolism, CSF tau levels begin to be elevated and become abnormal progressively. Grey matter volume changes, which become abnormal at last, can be detected by structural MRI which reflects brain atrophy and massive neuronal loss. Changes of each biomarker start from the follow a sigmoid shape which implies a rapid progression in the middle period and little change in initial and final periods (Clifford R. Jack *et al.*, 2010; Ewers *et al.*, 2011; Sperling *et al.*, 2011).

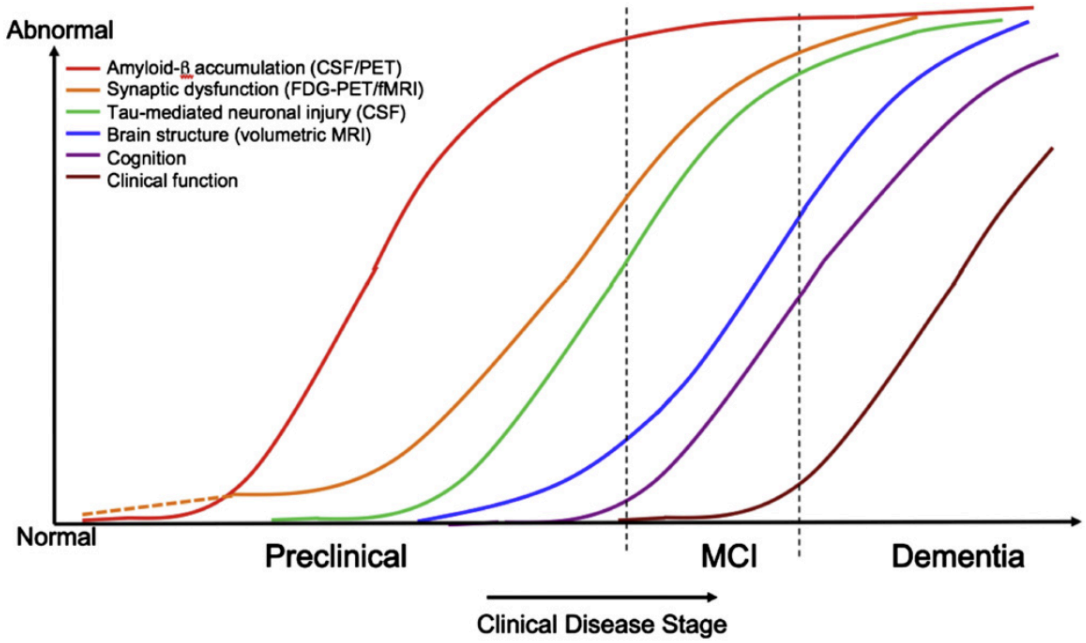

Figure 1. Hypothetical model of dynamic biomarkers of the AD expanded to explicate the preclinical phase. Aβ is identified by CSF Aβ42 or PET amyloid imaging (red). Synaptic dysfunction is assessed by FDG-PET or fMRI (orange). Tau is identified by CSF (green). Brain structure is evidenced by structural MR(blue) (adopted from (Sperling et al., 2011)).

For SAD, the individuals who will develop AD dementia cannot be confirmed until obvious disease characteristics being detected (Clifford R. Jack *et al.*, 2010). Because of the predictable parental onset age of ADAD patients, their pathophysiological changes can be assessed in relation to the estimated years to symptom onset (EYO).

The Dominant Inherited Alzheimer's Network (DIAN) is a multicenter study for the longitudinal assessment of neuroimaging, biomarker, and neuropsychological changes in ADAD (http://www.dian-info.org/de/). Levels of Aβ42 assessed in CSF or by amyloid PET becoming abnormal as early as 25 years before their estimated time point of symptom onset. Pathologic tau as assessed by CSF tau phosphorylated at threonine 181 ($p\text{-}tau_{181}$) shows a significant increase approximately 15 years before the expected age at symptom onset in ADAD. Besides pathological changes, structural changes such as hippocampal atrophy were observed 15 years before symptom onset, followed by cerebral glucose hypometabolism and cognitive impairment (measured by the Clinical Dementia Rating-Sum of Boxes (CDR-SB))

approximately 10 years before expected symptom onset (Bateman *et al.*, 2012). Temporal ordering of abnormalities in biomarkers is similar to the change during the course of SAD which suggested common pathophysiology between SAD and ADAD (Jack *et al.*, 2013; Villemagne *et al.*, 2013). The progress of biomarker change in SAD is similar compared to ADAD, but the symptoms are less severe and the onset is later. Similar to ADAD, CSF $A\beta_{1-42}$ also decreases in SAD but the degree of decline is only half of that in ADAD (Ringman *et al.*, 2012). The ratio of $A\beta_{1-40}$/ $A\beta_{1-42}$ is observed decrease in the progress of ADAD, however, in SAD, the reported results can be variable (Mayeux *et al.*, 2003; van Oijen *et al.*, 2006; Blennow *et al.*, 2010). For the tau pathology, higher severity has also been showen in ADAD compared to SAD (Sunderland *et al.*, 2003).

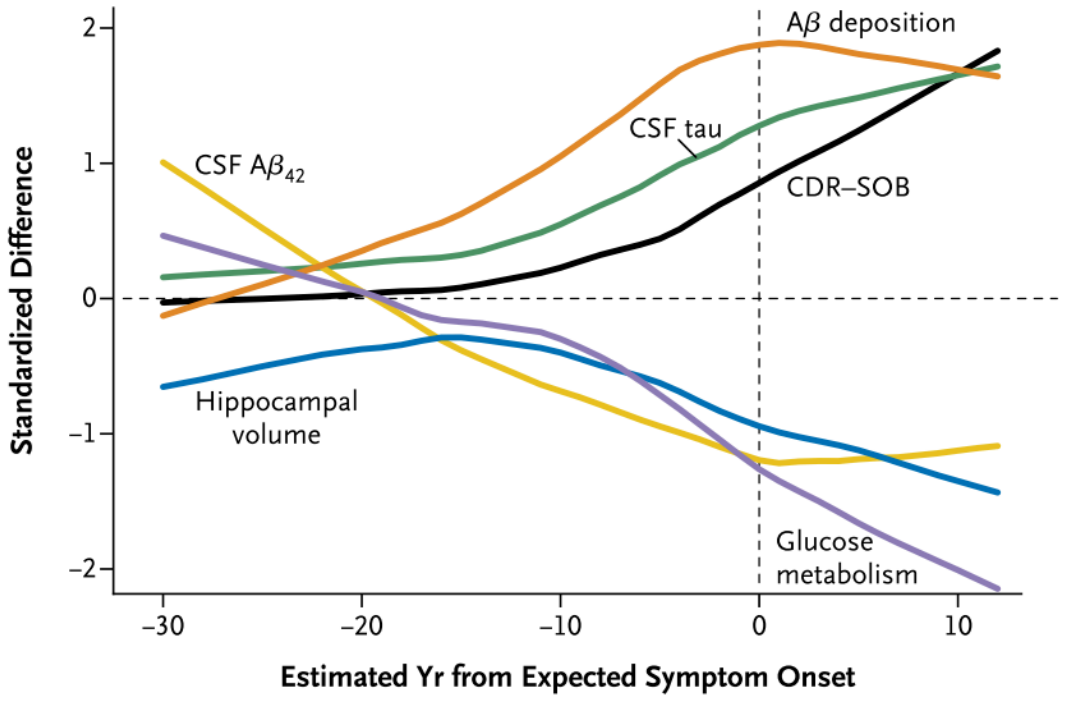

Figure 2. Temporal ordering of dynamic biomarkers in ADAD. Temporal ordering is assessed based on estimated years from expected symptom onset and shown with standardized difference between mutation and non-mutation carriers. The abnormalities of biomarkers appear in the following order: CSF Aβ42, fibrillar Aβ deposition, CSF tau, hippocampal atrophy, glucose hypometabolism, followed by cognitive change. (adopted from (Bateman *et al.*, 2012) *).*

1.3. Machine learning for brain imaging and bioinformatics

In recent years, the development of machine learning was growing very fast in many fields. Neuroscience is one of the fields with the fastest growth in machine learning applications (Vogt, 2018; Vu *et al.*, 2018). In the past decade, the number of publications about machine learning in neuroscience and AD increased rapidly (*Figure 3*).

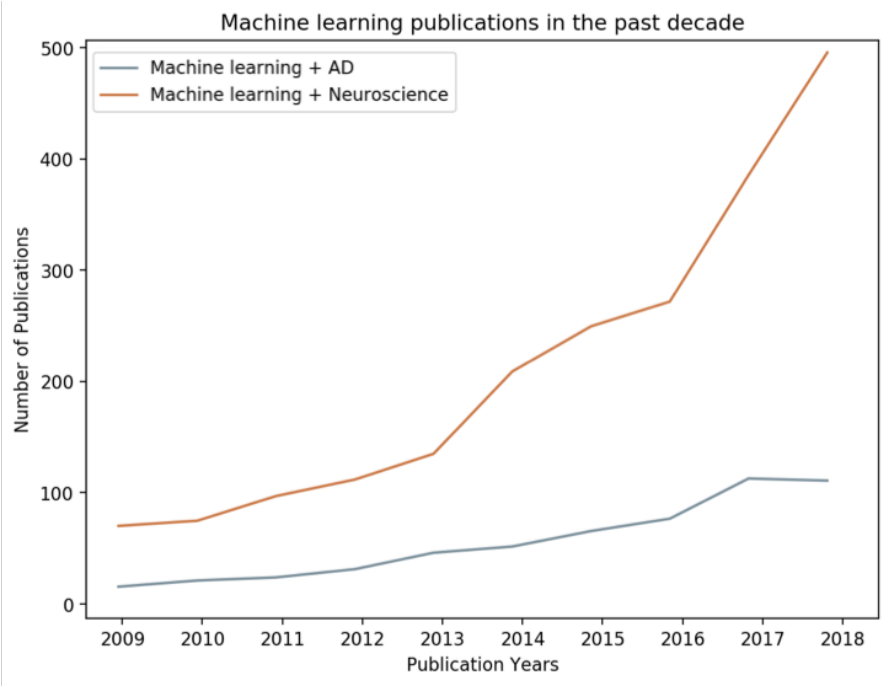

Figure 3. Trends of machine learning publications in neuroscience and Alzheimer's disease in the past decade. Numbers of publications from Web of Science based on search of keywords: machine learning Alzheimer's disease, machine learning neuroscience.

For volumetric MRI derived features, there are mainly two types of features: voxel-based features and region of interest (ROI) based features. Recent machine learning studies investigating the use of structural MRI to predict the conversion in the early stage of AD (MCI converters vs. non-converters) are listed in *Table 1*. Voxel-based features yielded AUC/accuracy of 70-86%. ROI-based features showed AUC/accuracy of 71-93%. Hippocampus features had the AUC/accuracy of 74-79% in distinguishing convertors with non-converters.

Table 1: Structural MRI as predictor: results from recent (past 5 years) studies on the classification of MCI who developed AD dementia (MCI converter) vs those who remained stable during the follow up period (MCI non-converter)

Study	**Dataset**	**Sample size: MCI cov: MCI ncov**	**AUC**	**SEN**	**SPE**	**Follow-up duration (month)**
Voxel-based features						
(Moradi *et al.*, 2015)	ADNI	164:100	76.61	88.85	51.59	0-36
(Wang *et al.*, 2016)	ADNI	64:65	69.7*	64.06	75.38	0-36
(Beheshti, Demirel and Matsuda, 2017)	ADNI	71:65	76.92	73.23	75.08	0-36
(Lin *et al.*, 2018)	ADNI	169:139	86.1	84	74.8	0-36
ROI-based features						
(Guerrero *et al.*, 2014)	ADNI	116:114	71*	75.0	67.0	0-24
(Chen, Wei and Liu, 2015)	ADNI	167:236	71.8	58.1	76.3	0-24
(Cheng *et al.*, 2015)	ADNI	43:56	76.40	74.3	72.1	0-24
(Clark *et al.*, 2016)	ADNI	24:83	76.0	68.2	75.6	0-48
(Hor and Moradi, 2016)	ADNI	96:126	84.8	81.9	75.0	0-36
(Korolev, Symonds and Bozoki, 2016)	ADNI	120:139	76.0	68.5	69.6	0-36
(Liu *et al.*, 2016)	ADNI	117:117	80.90	85.95	78.41	0-18
(Long *et al.*, 2017)	ADNI	95:132	93.2	86.32	90.91	0-36
Hippocampus						
(Komlagan *et al.*, 2014)	ADNI	166:236	75.6*	61.5	85.6	-
(Hu *et al.*, 2015)	ADNI	71:62	79.00	71.83	83.26	0-36
(Sørensen *et al.*, 2016)	ADNI	93:140	74.2	-	-	0-24

Abbreviations: ADNI = Alzheimer's Disease Neuroimaging Initiative; AUC = Area Under Curve; SEN = Sensivity; SPE = Specificity

* accuracy

Voxel-based methods quantify the brain structure changes by density map which were used in several studies in recent years (Moradi *et al.*, 2015; Wang *et al.*, 2016; Beheshti, Demirel and Matsuda, 2017; Lin *et al.*, 2018). ROI-based methods are based on pre-defined anatomical atlases such as Desikan-Killiany atlas (Desikan *et al.*, 2006) and automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002). The quantitative measures used in ROI-based methods include cortical thickness (Clark *et al.*, 2016; Korolev, Symonds and Bozoki,

2016), regional volume (Chen, Wei and Liu, 2015; Cheng *et al.*, 2015; Clark *et al.*, 2016; Hor and Moradi, 2016; Korolev, Symonds and Bozoki, 2016), and average intensity value (Guerrero *et al.*, 2014; Liu *et al.*, 2016; Long *et al.*, 2017). In both voxel or ROI based methods, some studies focused on the hippocampus (Komlagan *et al.*, 2014; Hu *et al.*, 2015; Sørensen *et al.*, 2016) and entorhinal cortex (Dickerson *et al.*, 2001). The voxel-based method achieved prediction accuracy or AUC of 76-85%. The ROI-based features were 57-82% accurate in differentiating MCI converters and non-converters.

Table 2: FDG-PET as predictor: results from recent (past 5 years) studies on the classification of MCI who developed AD dementia (MCI converter) vs those who remained stable during the follow up period (MCI nonconverter)

Study	**Dataset**	**Sample size: MCI cov: MCI ncov**	**AUC**	**SEN**	**SPE**	**Follow-up duration (month)**
Voxel-based features						
(Cabral *et al.*, 2015)	ADNI	44:56	85*	-	-	0-24
(Dukart, Sambataro and Bertolino, 2015)	ADNI	29:135	82.4	90.0	83.9	2-10 years
(Wang *et al.*, 2016)	ADNI	64:65	75.97*	68.74	83.08	0-36
ROI-based features						
(Cheng *et al.*, 2015)	ADNI	43:56	74.1	76.4	67.9	0-24
(Jie *et al.*, 2015)	ADNI	242:174	57.00	48.37	59.11	0-36
(Xu *et al.*, 2015)	ADNI	27:83	74.1	67.4	66.7	0-36
(Suk, Lee and Shen, 2015)	ADNI	43:56	68.9*	-	-	0-18
(Choi and Jin, 2018)	ADNI	79:92	82	70.9	79.3	0-36

Abbreviations: ADNI = Alzheimer's Disease Neuroimaging Initiative; AUC = Area Under Curve; SEN = Sensivity; SPE = Specificity

* accuracy

For FDG-PET, - as for structural MRI - there are two types of features for classification with FDG-PET: voxel-based features and ROI or atlas-based features. Therefore, features of FDG-PET studies for AD prediction can be roughly divided into voxel-based FDG-PET (Cabral *et al.*, 2015; Dukart, Sambataro and Bertolino, 2015) (Wang *et al.*, 2016), and ROI-based (Cheng *et al.*, 2015; Jie *et al.*, 2015; Suk, Lee and Shen, 2015; Choi and Jin, 2018) (Xu *et al.*, 2015) FDG-PET studies (*Table 2*).

Neuropsychological assessments have identified the cognitive and behavior changes in early AD. For cognitive alterations as a predictor for the identification of subjects who will convert to AD dementia, there were several studies using test such as verbal cued recall and verbal fluency tests, ADAS-Cog test, Functional Activities Questionnaire (FAQ), Rey Auditory Verbal Learning Test (RAVLT), Mini Mental State Examination (MMSE) and Geriatric Depression Scale (GDS) (Segovia *et al.*, 2014; Dukart, Sambataro and Bertolino, 2015; Clark *et al.*, 2016; Korolev, Symonds and Bozoki, 2016; Wang *et al.*, 2016) (*Table 3*). With neuropsychological measures, the studies showed AUC/accuracy of 72-87%.

Table 3 Neuropsychological scores as predictors: results from recent (past 5 years) studies on the classification of MCI who developed AD dementia (MCI converter) vs those who remained stable during the follow up period (MCI non-converter)

Study	**Dataset**	**PSY tests**	**Sample size: MCI cov: MCI ncov**	**AUC**	**SEN**	**SPE**	**Follow-up duration (month)**
(Segovia *et al.*, 2014)	EADC	Verbal cued recall, Verbal fluency	26:20	73.91*	73.08	75.00	0-36
(Dukart, Sambataro and Bertolino, 2015)	ADNI	MMSE, GDS, ADAS, RAVLT, FAQ	29:135	71.6	85.7	51.3	0-24
(Clark *et al.*, 2016)	ADNI	Verbal fluency	24:83	87.2	70.8	88	0-48
(Korolev, Symonds and Bozoki, 2016)	ADNI	ADAS-Cog, FAQ, RAVLT	120:139	83.0	76.9	75.3	0-36
(Wang *et al.*, 2016)	ADNI	ADAS-Cog	64:65	79.07*	73.44	84.62	0-36

Abbreviations: EADC =European Alzheimer's disease consortium; ADNI = Alzheimer's Disease Neuroimaging Initiative; AUC = Area Under Curve; SEN = Sensivity; SPE = Specificity; MMSE = Mini Mental State Examination; GDS = Geriatric Depression Scale; ADAS = Alzheimer's Disease Assessment Score; RAVLT = Rey Auditory Verbal Learning Test

* accuracy

Previous studies established CSF biomarkers Aβ42, t-tau, and p-tau to predict MCI conversion to AD as single modality (*Table 4*). For the classification of MCI converters versus non-converter, the results CSF biomarkers were between 58% to 68% (AUC/accuracy) when used as a single marker.

Table 4: CSF biomarkers as predictor: results from recent (past 5 years) studies on the classification of MCI who developed AD dementia (MCI converter) vs those who remained stable during the follow up period (MCI nonconverter)

Study	**Dataset**	**Sample size: MCI cov: MCI ncov**	**AUC**	**SEN**	**SPE**	**Follow-up duration (month)**
		$A\beta_{1-42}$, p-tau, total Tau				
(Cheng *et al.*, 2015)	ADNI	43:56	67.6	74.6	61.5	0-24
(Jie *et al.*, 2015)	ADNI	242:174	63.00	53.02	63.04	0-36
(Suk, Lee and Shen, 2015)	ADNI	43:56	57.7*	-	-	0-18

Abbreviations: ADNI = Alzheimer's Disease Neuroimaging Initiative; AUC = Area Under Curve; SEN = Sensivity; SPE = Specificity

* accuracy

To predict the conversion in the early stage of AD, combining multiple modalities may provide complementary information and then improve the classification performance compared to a single modality. Some recent studies combined two modalities such as MRI & FDG-PET (Liu *et al.*, 2014; Suk, Lee and Shen, 2014; Hor and Moradi, 2016), MRI & neuropsychological tests (Korolev, Symonds and Bozoki, 2016; Minhas *et al.*, 2017), MRI & CSF biomarkers (Frölich *et al.*, 2017). Some other multi-model studies did the classification combined with three modalities such as MRI & FDG-PET & neuropsychological tests (Segovia *et al.*, 2014; Dukart, Sambataro and Bertolino, 2015; Wang *et al.*, 2016), MRI & FDG-PET & CSF biomarkers (Suk, Lee and Shen, 2014; Zhu, Suk and Shen, 2014; Cheng *et al.*, 2015; Jie *et al.*, 2015), MRI & FDG-PET & ^{18}F-florbetapir-PET (Xu *et al.*, 2015), MRI & neuropsychological tests & demographic data (Moradi *et al.*, 2015; Clark *et al.*, 2016; Tong *et al.*, 2017). The two-modality-combination achieved prediction accuracy or AUC of 71-87% (*Table 5*). With three-modality-combination, the studies showed the AUC/accuracy of 74-90% (*Table 5*).

Table 5: Multi-modal predictors: results from recent (past 5 years) studies on the classification of MCI who developed AD dementia (MCI converter) vs those who remained stable during the follow up period (MCI non-converter)

Study	**Dataset**	**Modalities**	**Sample size: MCI cov: MCI ncov**	**AUC**	**SEN**	**SPE**	**Follow-up duration (month)**
2-modality-combination							
(Liu *et al.*, 2014)	ADNI	MRI+FDG	43:56	69.57	64.88	70.00	0-18
(Suk, Lee and Shen, 2014)	ADNI	MRI+FDG	76:128	74.66	48.04	95.23	-
(Hor and Moradi, 2016)	ADNI	MRI+FDG	27:144	87.2	83.10	80.3	0-36
(Korolev, Symonds and Bozoki, 2016)	ADNI	MRI+PSY	120:139	87.0	83.4	76.4	0-36
(Minhas *et al.*, 2017)	ADNI	MRI+PSY	67:78	-	92.3	87.5	0-36
(Frölich *et al.*, 2017)	DCN	MRI+CSF	28:87	82	85	64	0-36
3-modality-combination							
(Segovia *et al.*, 2014)	EADC	MRI+FDG+PSY	26:20	86.96*	92.32	80.00	0-36
(Dukart, Sambataro and Bertolino, 2015)	ADNI	MRI+FDG+PSY	29:135	83.3	100	75.5	0-24
(Wang *et al.*, 2016)	ADNI	MRI+FDG+PSY	64:65	84.5*	82.81	86.15	0-36
(Jie *et al.*, 2015)	ADNI	MRI+FDG+CSF	242:174	72.00	66.05	76.61	0-36
(Zhu, Suk and Shen, 2014)	ADNI	MRI+FDG+CSF	43:56	78.8	48.5	94.4	0-36
(Suk, Lee and Shen, 2015)	ADNI	MRI+FDG+CSF	43:56	83.3*	-	-	0-18
(Cheng *et al.*, 2015)	ADNI	MRI+FDG+CSF	43:56	84.80	84.50	72.70	0-24
(Moradi *et al.*, 2015)	ADNI	MRI+PSY+AGE	164:100	90.20	86.65	73.64	0-36
(Tong *et al.*, 2017)	ADNI	MRI+PSY+AGE	171:129	87.0	86.7	72.6	0-36
(Clark *et al.*, 2016)	ADNI	MRI+PSY+DEM	24:83	81.4	62.5	89.2	0-48
(Xu *et al.*, 2015)	ADNI	MRI+FDG+FBP	27:83	80.7	74.1	81.5	0-36

Abbreviations: EADC =European Alzheimer's disease consortium; ADNI = Alzheimer's Disease Neuroimaging Initiative; DCN = Dementia Competence Network; AUC = Area Under Curve; SEN = Sensivity; SPE = Specificity; FBP= florbetapir; DEM = Demographic Variable

* accuracy

From the articles which focused on automated detection of MCI conversion to AD in the past five years, the majority of studies combined multiple modalities rather than single modality (Some of the single modality results listed in *Table 1-4* are parts of the multimodality studies listed in *Table 5*). Based on the use of imaging data – MRI, FDG-PET or the combination of those, the classification accuracy is generally below 80%. Only two of five articles listed in *Table 3* which used neuropsychological tests showed a classification accuracy higher than 80%. Using only CSF biomarkers, the classification accuracy was in none of the studies above 70%. Compared to the single modality, multi-modality predictors achieved better performance which was generally above 80% classification accuracy.

Machine learning can be roughly categorized into supervised learning, unsupervised learning and semi-supervised learning. In supervised learning, the prediction model is established based on the data which is associated with the output and then applied to the new data for prediction (such as classification). In unsupervised learning refers to the models are trained based on the unlabelled data. In semi-supervised learning, the models are established based on both labeled and unlabeled data. Its purpose is to use limited labeled data and a large amount of unlabeled data to enhance the learning ability.

To improve classification performance, feature selection is a critical step by selecting the most discriminative features. Among the supervised methods, Information Gain (IG) (Hor and Moradi, 2016), t-test (Xu *et al.*, 2015; Beheshti, Demirel and Matsuda, 2017; Minhas *et al.*, 2017), Mutual Information (MI) (Cabral *et al.*, 2015), and Markov Blanket approach (Dukart, Sambataro and Bertolino, 2015) were used in recent five years' studies of predicting AD conversion. When the number of features is very high and exceeds the number of training samples such as voxel-based features of imaging data, it will lead to poor performance of classification. To reduce the feature dimensionality, unsupervised learning methods were usually applied including Principle Component Analysis (PCA) (Wang *et al.*, 2016; Choi and Jin, 2018; Lin *et al.*, 2018), Independent Component Analysis (ICA) (Segovia *et al.*,

2014), Least Absolution Shrinkage and Selection Operator (LASSO) (Guerrero *et al.*, 2014; Zhu, Suk and Shen, 2014; Tong *et al.*, 2017; Lin *et al.*, 2018). For classification, a number of algorithms on predicting MCI conversion to AD have been used based on supervised and semi-supervised learning. The most used supervised learning method in recent research is Support Vector Machine (SVM) (Liu *et al.*, 2014, 2016; Segovia *et al.*, 2014; Cabral *et al.*, 2015; Cheng *et al.*, 2015; Hu *et al.*, 2015; Jie *et al.*, 2015; Sørensen *et al.*, 2016; Beheshti, Demirel and Matsuda, 2017). Besides, Naïve Bayes (NB) (Cabral *et al.*, 2015; Dukart, Sambataro and Bertolino, 2015), Random Forest (RF) (Hor and Moradi, 2016; Tong *et al.*, 2017), and logistic regression (Clark *et al.*, 2016) were also widely used supervised learning methods in predicting AD. With semi-supervised methods, few recent studies investigated the used of Low Density Separation (LDS) (Moradi *et al.*, 2015) and multimodal relevance vector regression (Zhu, Suk and Shen, 2014) for the predicting of MCI conversion.

1.4. Aim of the study

The overall goal of the current thesis was to establish a cross-validated multimodal biomarker model for the early detection of AD. The specific aims were

1) Train the best classification model based on single modalities and combinations of up to 4 modalities for discriminating ADAD against controls.

2) Cross-validate the best classification models in a large sample of SAD subjects at different disease stages

2 Materials and methods

2.1. Databases

DIAN:

Dominantly Inherited Alzheimer's Network is a multisite study established in 2008 (Moulder *et al.*, 2013). The DIAN study recruits individuals who carry the gene mutations (PSEN1, PSEN2 or APP) and their siblings who do not carry any of the mutations. The individuals who are non-mutation carriers contribute as control sample. The ratio of mutation and non-mutation carriers is approximately equal. Inclusion criteria are: 1) provide informed consent before other procedures; 2) older than 18 years; 3) nursing home-level care is not needed; 4) have at least one non-full-blooded sibling as collateral source; 5) language fluency needs to be equal or higher than 6^{th} grade level. The participants who have psychiatric illness were not eligible. The main goals of DIAN study are to investigate asymptomatic brain changes with cognitive, imaging and fluid biomarkers to track the disease.

ADNI:

Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multicenter study that was established by various institutes and organizations (e.g. National Institute on Aging (NIA)). ADNI started in October 2004 led by Dr. Michael W. Weiner from department of Radiology and Biomedical Imaging in University of California, San Francisco (UCSF). The first stage was ADNI-1 which has recruited around 800 subjects (HC/MCI/AD = 200/400/200) for five years. Different types of data were collected in ADNI-1 including brain scans (structural MRI, FDG-PET, PIB-PET), CSF biomarkers ($A\beta_{1\text{-}42}$, p-tau, total Tau) and genetic profiles. The initial aim was developing biomarkers for tracking the early stage of AD. The second stage ADNI-GO started in September 2009 and recruited another 200 new early MCI subjects. At the same time, the subjects from ADNI-1 were keeping examined. Besides the

data types measured ADNI-1, diffusion MRI (diffusion tensor imaging, DTI) and resting state fMRI were added in ADNI-GO. The focus of this stage is biomarkers of earlier stage of AD. Followed ADNI-GO, the third stage ADNI-2 begun in September 2011 and lasted for five years. In addition to follow up the participants of ADNI-1/GO, 150 HC, 300 MCI and 150 AD subjects were added. Main differences from previous stages were additional amyloid PET (florbetapir PET) added. The past three stages ADNI-1/GO/2 recruited more than 1500 subjects from 55 to 90 years old and widely used in clinical and scientific research.

2.2. Participants

DIAN:

A total of 174 subjects were included from DIAN cohort who were enrolled from September 2009 through April 2014 collected at 13 sites (7 sites in the United States, 3 sites in Australia, 1 site in the United Kingdom, 2 sites in Germany). Of these subjects, 101 were carriers of mutation causing AD in the genes presenilin1 (PSEN1, n = 77), presenilin2 (PSEN2, n = 9), or amyloid precursor protein (APP, n = 15). And 73 subjects were non-mutation carriers from the same families that were used as control samples (*Table 6*). In addition to satisfy the inclusion criteria of DIAN mentioned above, all DIAN subjects included in this study had to be assessed with T1-weighted MRI, FDG-PET, neuropsychological tests and CSF biomarkers of $A\beta_{1-42}$ and tau which is shown in *Figure 4A*. For each subject, the predictable year to symptom onset was estimated based on the symptomatic onset of parents. EYO thereby was defined as current age minus the year of symptom onset.

ADNI:

We included 545 subjects with baseline diagnoses from ADNI GO/2 database as our validation sample set. Among all the subjects, there are 205 HC subjects, 262 MCI subjects, and 78 AD patients (*Table 6*). HC refers to the subjects of cognitive normal with mini-mental

state examination (MMSE) score between 24 and 30, clinical dementia rating (CDR) of 0 and no signs of cognitive impairment and significant depression. MCI indicates the subjects who have mild cognitive impairment with MMSE score between 24 and 30, CDR score of 0.5 and subjective memory concern. AD subjects had MMSE score between 20 and 26, a CDR score equal or greater than 0.5 and memory complaints. In addition to meet these requirements, ADNI inclusion criteria mainly included: Geriatric Depression Scale < 6, age from 55 to 90, fluently speaking English or Spanish, willing to attend longitudinal measures (more details in ADNI protocol). The ADNI subjects involved in our study also need to have 3T T1-structural MRI, FDG-PET, neuropsychological tests and CSF biomarkers at baseline and at least one clinical follow-up assessment after 3 years (*Figure 4B)*. HC converters were defined based on conversion MCI or AD dementia within 36 months. MCI converters were defined based on conversion to AD dementia within 36 months. The other HC or MCI subjects who didn't convert and had at least 36 months follow-up data were included as non-converters. Median clinical follow-up time was 48 months for HC subjects (from 12 up to 60 months) and also 48 months for MCI subjects (from 36 up to 60 months).

Table 6A: Baseline demographics of DIAN

Demographic	**Mutation Carriers**	**Non-carriers**
n	101	73
Age[a]	39.9(10.4)	39.2(10.5)
Male(%)	48(48)	33(45)
Education[a]	13.7(3.0)**	14.9(2.3)
APOE4+(%)	59(58.4)***	2(2.7)
ADAS11[a]	26(26.0)	23(32.4)
MMSE[a]	26.6(9.1)*	29.0 (1.3)

Table 7B: Baseline demographics of ADNI

Demographic	HC	AD	HC cov	HC ncov	MCI cov	MCI ncov
n	205	78	14	62	78	184
Age[a]	72.7(6.1)**	75.6(7.7)	76.3(5.9)	74.1(6.0)	72.9(7.0)	71.1(7.2)
Male(%)	110(46)	60(61)	11(61)	37(50)	42(52)	116(56)
Education[a]	16.6(2.5)**	15.6(2.6)	15.6(2.7)	16.6(2.6)	16.1(2.6)	16.3(2.6)
APOE4+(%)	35(29)***	67(69)	4(22)	21(28)	60(74)**	80(38)
ADAS11[a]	5.6(2.9)***	19.7(6.4)	8.7(2.8)*	6.0(2.9)	13.5(4.8)**	8.1(3.3)
MMSE[a]	29.0(1.2)***	22.9(2.0)	28.6(1.4)	29.1(1.4)	27.2(1.7)**	29.1(1.4)

Cov = converter, ncov = non-converter
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$
[a] values are mean and standard deviation

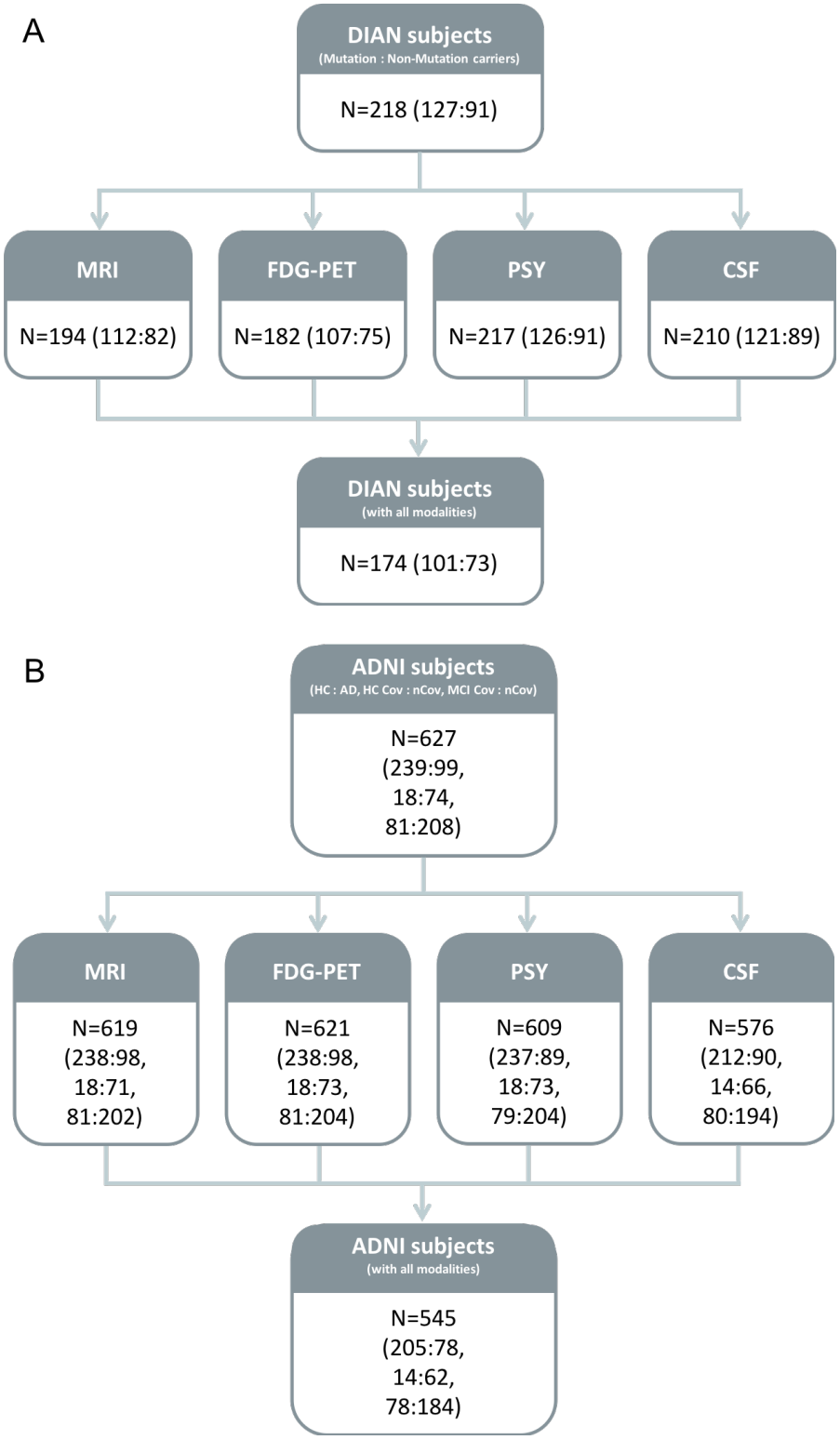

Figure 4. Flowchart for subject inclusion. In (A) DIAN and (B) ADNI, we included the subjects who had the data of all four modalities (MRI, FDG-PET, psychological data and CSF biomarkers). Abbreviation: Cov = converter, nCov = non-converter

2.3. MRI acquisition and assessment

DIAN:

For each DIAN subject, structural MRI was acquired with Siemens 3T scanner and pass the initial and follow-up quality control procedures. Across different sites, scans were performed based on unified protocol (according to ADNI protocol). Three-dimensional T1-weighted sagittal MP-RAGE scans lasted for five minutes with the following parameters: TR/TE=2300ms/2.95ms, flip angle = 9°, 1.1 × 1.1 × 1.2 mm resolution. All the scans were screened by DIAN Imaging Core for compliance of the imaging acquisition protocol. Processing of 3T MRI scans was done with FreeSurfer 5.1 (FreeSurfer Software Suite is available at http:// surfer.nmr.mgh.harvard.edu/) based on Dell PowerEdge 1950 servers with Intel Xeon processors contained cortical surface reconstruction and volumetric segmentation (Fischl, 2012). DIAN MRI processing included motion correction, segmentation of volumetric structure (gray matter and white matter) (Fischl *et al.*, 2002), intensity normalization, cortical surface extraction and parcellation of cortical and subcortical areas using Desikan Killiany probabilistic atlas (Desikan *et al.*, 2006). Each region of interest (ROI) was corrected for intracranial volume with published method (Jack *et al.*, 1989). All MRI processing was done by the DIAN imaging core including ROI extraction. For further analysis, we used cortical volume from 41 ROIs for each hemisphere (34 cortical and 7 subcortical ROIs) (*Figure 5*). All Desikan Killiany ROIs are listed in *Table 8*.

ADNI:

T1 MRI images of ADNI database were acquired using 3T Siemens scanners. All scans were based on ADNI acquisition protocol and posted methods (available at http://www.adni-info.org). Imaging was performed with three-dimensional MP-RAGE sequence with the following parameters: TR/TE/TI=2300/2.98/900ms, field of view (FOV) = 256 × 240 mm^2, flip angle = 9°, 176 slices, 1.1 × 1.1 × 1.2 mm resolution. Each MP-RAGE image underwent

gradient non-linearity distortion correction, B1 non-uniformity correction and non-uniform intensity normalization (N3) by ADNI MRI core (http://adni.loni.usc.edu/methods/mri-tool/mri-pre-processing/).

The volumetric T1-weighted images were processed using Freesurfer (Version 5.1). The procedure done by ADNI Imaging Core included motion correction, geometry distortion correction, non-brain tissue removal, Talairach space transformation, white matter and gray matter segmentation and intensity normalization. Desikan-Killiany atlas-based ROI extraction was also provided by ADNI Core. We used same extracted ROIs (41 ROIs for each hemisphere) subdivided from cortical and subcortical regions of ADNI MR scans as DIAN data for further analysis.

2.4. FDG-PET acquisition and assessment

DIAN:

^{18}F-FDG scans were performed using a bolus injection of 5 mCi of FDG. The acquisition started 40 minutes after the injection and lasted for 20 minutes (4x5 minute frames). Each FDG-PET image was taken quality check according to the ADNI PET QC. With the standard protocol, PET images were motion-corrected and registered to individual FreeSurfer derived MRI regions of interest. FDG-PET was then converted to standardized uptake value ratio (SUVR) for each ROI in Desikan-Killany Atlas space with the brainstem reference. Due to lower spatial resolution of PET imaging than structural MR imaging, it will lead to loss of activity or blurring of adjacent tissues which is called partial volume effects (PVE). So partial volume correction (PVC) was done using a method based on regional spread function (RSF) by DIAN imaging core (Roussel, Ma and Evans, 1998).

ADNI:

Before PET scanning, each subject need to be checked to ensure blood glucose <180 mg/dL. FDG-PET scanning will last for 30 minutes with 4x5 minute frames after 30 minutes of injection of 185 MBq (5 mCi) fluorodeoxyglucose. After scanning, each PET image was checked by ADNI PET QC team. In order to get uniform PET scans, same Freesurfer preprocessing pipeline (as described previously in DIAN) was applied to ADNI dataset. Since FDG-PET SUVR value is not available from ADNI core, we performed the same procedure to calculate the corresponding SUVR score by superimposing Desikan-Killiany ROIs to FDG-PET. SUVR value of ADNI PET images was normalized to brainstem.

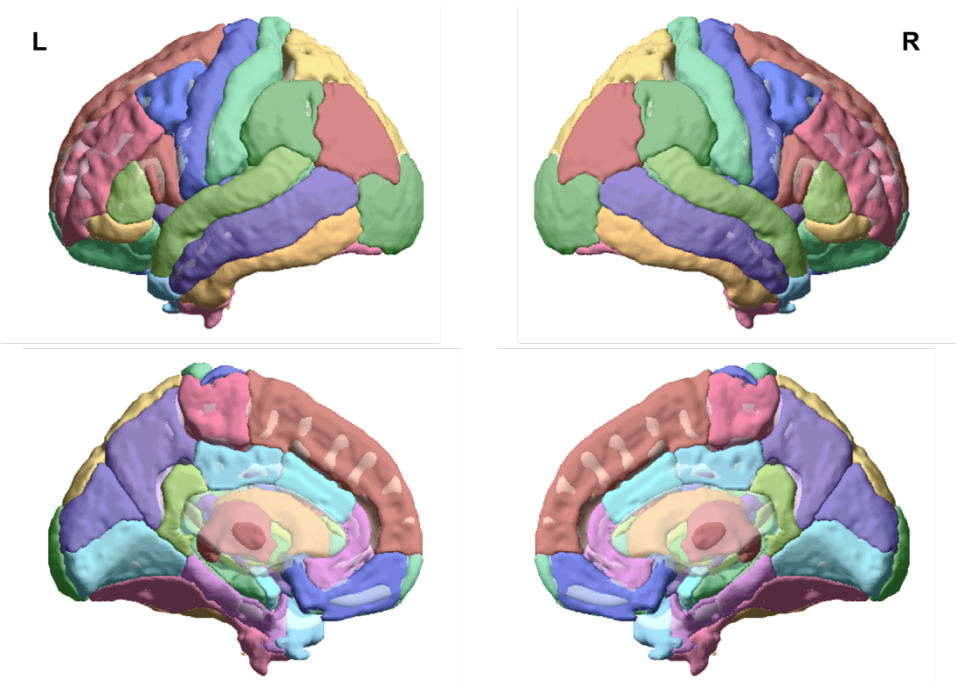

Figure 5. Regions of Desikan-Killiany atlas. For each hemisphere, it includes 34 cortical regions and 7 subcortical regions.

Table 8: Regions of Desikan-Killiany atlas

Frontal lobe	***Temporal lobe***
Caudal middle frontal gyrus	Banks superior temporal sulcus
Frontal pole	Inferior temporal gyrus
Lateral orbital frontal cortex	Middle temporal gyrus
Medial orbital frontal cortex	Superior temporal gyrus
Paracentral lobule	Transverse temporal cortex
Pars opercularis	Entorhinal cortex
Pars orbitalis	Fusiform gyrus
Pars triangularis	Parahippocampal gyrus
Precentral gyrus	Temporal pole
Rostral middle frontal gyrus	***Cingulate cortex***
Superior frontal gyrus	Caudal anterior-cingulate cortex
Parietal lobe	Insula
Inferior parietal cortex	Isthmus–cingulate cortex
Postcentral gyrus	Posterior-cingulate cortex
Precuneus cortex	Rostral anterior cingulate cortex
Superior parietal cortex	***Subcorical structures***
Supramarginal gyrus	accumbens
Occipital lobe	amygdala
Cuneus cortex	caudate
Lateral occipital cortex	hippocampus
Lingual gyrus	pallidum
Pericalcarine cortex	putamen
	thalamus

2.5. Neuropsychological tests

Participants from DIAN and ADNI underwent a set of neuropsychological assessments. Among these assessments, we included seven tests for both DIAN and ADNI studies which are listed below.

1) Mini-Mental State Examination (MMSE) is a widely used test to evaluate cognitive impairment and screen for possible dementia. MMSE score ranges from 0 to 30: greater than 23 points suggests cognitively normal; 19-23 indicates mild cognitive impairment; 10-18 means moderate cognitive impairment; below 10 suggests severe cognitive

dysfunction. The test evaluates orientation (time and place), registration (repetition immediately), attention and calculation, recall, language, repetition and visual construction (Folstein, Folstein and McHugh, 1975; Weintraub *et al.*, 2009).

2) Boston Naming Test is a measure of naming ability and semantic memory. During test, participants will be showed 30 pictures and asked to tell the name of the object. In both DIAN and ADNI studies, reduced version was used based on original 60 items test (Goodglass and Kaplan, 1983).
3) Category Fluency Test, also called verbal fluency test, is a measure of semantic memory and executive function. Participants will be asked to produce words of a given category within 60 seconds. We used category fluency test of animal category in this study from DIAN and ADNI (Morris *et al.*, 1989).
4) Trail Making Test, including two different tests (A&B), is a measure of processing speed visual–motor skills, and executive functions. For the test A, participants need to connect the 25 circles from number 1 through 25 in sequence. For the test B, there are also 25 circles including 13 numbers (1-13) and letters (A-L) which need to be connected in alternative order. Protocols are same for DIAN and ADNI (Reitan and Wolfson, 1985).
5) Word List Immediate and Delayed Recall test is a measure of episodic memory. Participants will be asked to recall a list of words immediately and after a while. In DIAN, the word list contains 16 high-frequency words. In ADNI, the word list contains 10 words (Rosen, Mohs and Davis, 1984).

2.6. CSF biomarkers

CSF concentration of amyloid-β_{42}, total tau and phosphorylated tau at threonine 181 (p-tau) were included in our analysis. The protocol of gathering CSF in DIAN is consistent with the biofluid protocol in ADNI.

DIAN:

CSF (15 ml) was collected in the morning under fasting circumstance by standard lumbar puncture into polypropylene tubes. CSF samples were then shipped to and analyzed by DIAN biomarker core team at Washington University. CSF amyloid-β_{42}, tau and p-Tau were measured by multiplex xMAP Luminex platform (Fagan *et al.*, 2014a).

ADNI:

CSF collection pipeline is the same with DIAN. CSF samples were shipped to and analyzed by the Roche Elecsys® electrochemiluminescence immunoassays and conducted by the ADNI biomarker core team at University of Pennsylvania (Shaw *et al.*, 2009).

2.7. Data preprocessing

We used DIAN data as our training data to establish the Naïve Bayes algorithm-based machine learning model and ADNI data as our independent validation data to test the classification model. To meet the prerequisites of classification, we did data preprocessing including normalization and standardization.

2.7.1 Feature normalization

Recent research shows Box-Cox transformation did improve the performance of classifiers which are under the assumption of normality (Bicego and Baldo, 2016). We used Naïve Bayes algorithm as our classification method. One of the prerequisites of Naïve Bayes with Gaussian kernel is that each attribute (feature) need to be normally distributed. Therefore, we did normalization as the first step of data preprocessing.

One of the most useful normalization methods is Box-Cox power transformation which was proposed by George Box and David Cox in 1964. The aim of Box-Cox transformation is to transform the non-normal data to approximately normal distributed. The reasons for non-normality usually are: 1) there are extreme values which may lead to a skewed distribution; 2)

limit of the data may also result in a skewed distribution; 3) overlap of more than one process will cause non-normal distribution. 4) insufficient data discrimination will make data become non-normal. To normalize the data, we used the following equation:

$$\mathrm{x}(\lambda) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0 \\ \ln x\,, & if\ \lambda = 0, \end{cases}$$

where x is the recorded data (features of different modalities) and x needs to be positive (>0). In our projects, some features didn't meet this condition: e.g. the minimum score of delayed word recall test is zero (in both DIAN and ADNI). If y is smaller or equal to zero, we used an extended form:

$$\mathrm{x}(\lambda) = \begin{cases} \dfrac{(x + c)^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0 \\ \ln(x + c)\,, & if\ \lambda = 0, \end{cases}$$

where c is a constant which makes any y meet the condition: $x + c > 0$. In both equations, λ is the transformation parameter estimated using maximum likelihood theory. The operation with different values of parameter λ is shown in *Figure 6*. In special cases, when $\lambda = 0$, it's natural log transformation. When $\lambda = 1$, it's linear transformation and the data stays unchanged. When $\lambda = -1$, it's reciprocal transformation. When $\lambda = 0$, it's natural log transformation. In our project, Box Cox transformation was done with SciPy library of Python (available at https://scipy.org/).

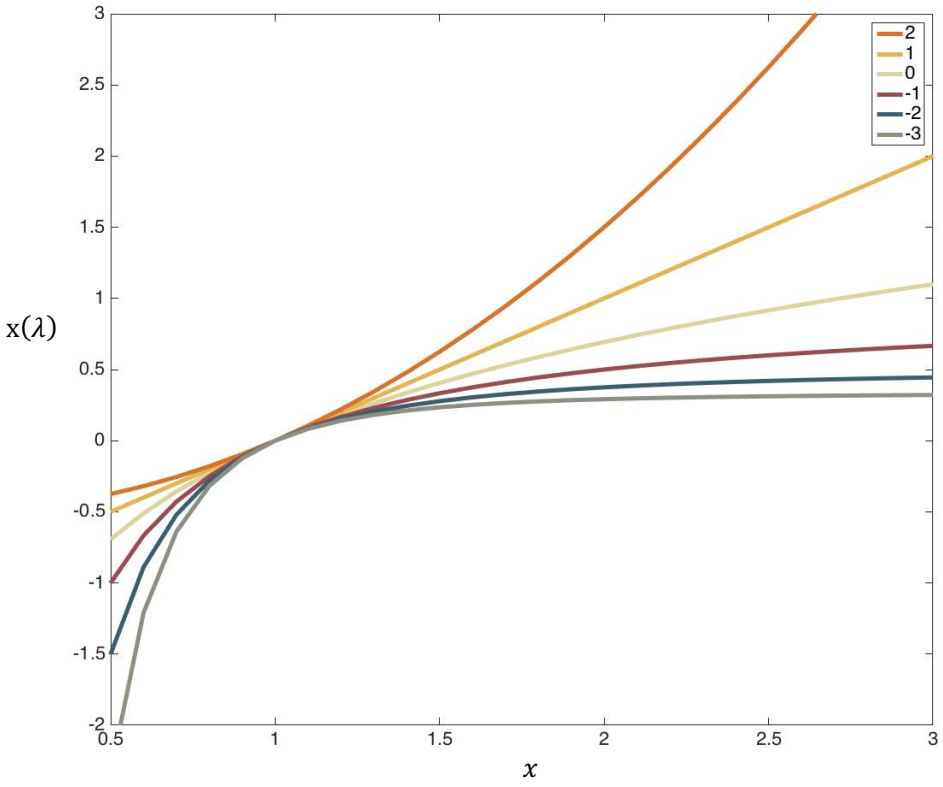

Figure 6. Box Cox transformation with different λ parameters

2.7.2 Feature standardization

A prerequisite for the cross-validation between studies is that the features across DIAN and ADNI are comparable. In other words, the same feature in DIAN and ADNI need to be in a comparable value range. Even though the assessment of DIAN and ADNI data used very similar protocols , some differences still remain (available at http://www.adni-info.org and dian.wustl.edu). Different CSF assays were used (xMap in DIAN and Elesys in ADNI) which may lead to difference in absolute CSF measurement values. Moreover, the age range of the subjects differs between the two studies (20-61 years old of mutation carriers in DIAN, 55-91years old of MCI in ADNI). Lack of consistency in training and validation set will lead to poor performance of classifier. Hence, we did standardization (scaling) as the second step of preprocessing.

We used the Centiloid method which was originally developed for the standardization of amyloid PET imaging and enabled the comparison of results from different tracers and methods (Klunk *et al.*, 2015). In this project, the Centiloid method was applied to create comparable value ranges (i.e. a scale ranging from 0 to100) across the training set (DIAN) and test set (ADNI) for each feature of MRI and FDG-PET images, psychological tests and CSF biomarkers. Similar to the procedure for the Centiloid scaling method applied to amyloid PET, we firstly defined two reference groups at both ends of the disease spectrum (HC vs AD dementia) for both DIAN and ADNI. The HC reference group included cognitive normal elder subjects who showed normal CSF Aß values. The AD reference group was included cases with AD dementia who showed abnormal CSF Aß values. Specifically, in DIAN, the control reference group was defined as non-mutation carriers who meet the conditions: EYO>0 & CDR=0 & Aß- (CSF $A\beta_{1\text{-}42}$ > 192 pg/ml). AD reference group was defined as mutation carriers who meet the conditions: EYO>0 & CDR>=1 & Aß+ (CSF $A\beta_{1\text{-}42}$ < 192 pg/ml) (Shaw *et al.*, 2009). In ADNI, the control reference group was defined as HC subjects who meet the conditions: Age<70 & Aß-. AD reference group was defined as AD subjects who met the conditions: Aß+. Limitation of age is to minimize age-related effects (non-Alzheimer related pathologies). Each feature was scaled linearly with the following equation (Jack *et al.*, 2015):

$$X_{scaled} = 100 \times \frac{X_{IND} - X_{Min}}{X_{Max} - X_{Min}}$$

Where:

X_{IND} is individual's feature value.

X_{Min} is the 5th percentile of $X_{Control}$ if $\overline{X_{Control}} < \overline{X_{AD}}$.

X_{Min} is the 5th percentile of X_{AD} if $\overline{X_{AD}} < \overline{X_{Control}}$.

X_{Max} is the 95th percentile of $X_{Control}$ if $\overline{X_{Control}} > \overline{X_{AD}}$.

X_{Max} is the 95th percentile of X_{AD} if $\overline{X_{AD}} > \overline{X_{Control}}$.

Scaled value was calculated for each feature and the comparison between original and scaled feature is shown in the chapter 3.

2.8. Feature selection

The total features we extracted include: 82 features from MRI (gray matter volume), 82 features from FDG-PET (SUVR), 7 features from neuropsychological tests and 3 CSF biomarker features. To enhance the performance of classification algorithm, we reduced the number of features and selected the most informative features using a method of information theory – information gain (IG) (Hall and Holmes, 2003; Quinlan, 2014). IG method, in our case, describes the amount of information gained about a feature from class separation. This method has successfully been used in previous studies on DTI and structural MRI for feature selection (Plant *et al.*, 2010; Dyrba *et al.*, 2013).

Information gain is based on Shannon entropy which is a measure of unpredictability in information theory. When a state is unpredictable, the entropy is high. In the contrast, when a state is predictable, the entropy is low. In our case, the entropy of class distribution is defined as $H(C) = -\sum_{c_i \in C} p(c_i) \cdot \log_2\big(p(c_i)\big)$. Where C is separated classes, c_i is the ith class and $p(c_i)$ is the probability of c_i. Class distribution here means the distribution of separated classes, i.e. if C = {AD, HC} and the number of subjects in HC and AD is equal, the class distribution is uniform distribution. If subjects in each class is balanced, the entropy is high ($H(C) = 1$). On the contrary, if it is unbalanced, the entropy is relatively low. The more unbalanced the class size is, the lower the entropy.

Based on entropy, information gain is defined as $IG(f_i) = H(C) - H(C|f_i)$, and $H(C|f_i)$ is the conditional entropy of the class distribution given the feature f_i. IG describes the decrease in entropy of the class distribution and conditional class distribution given feature f_i which indicates how much a feature contributes to the class separation. The range of IG value is

from 0 to 1. The larger the IG is, the more the information the feature gains for data differentiation.

For computing conditional entropy, the feature need to be discretized if it is continuous. We used an algorithm based on minimum description length principle (MDLP) invented by Fayyad and Irani which optimizes the number and location of cut points for each feature (Irani and Fayyad, 1993).

Since IG value of each feature indicates how it contributes to the class separation, to choose the most informative features, we only keep the features whose IG value is above the threshold for further classification. IG threshold varies for different modalities and different fold which will be described with more details in the next section.

2.9. Classification

Machine learning (ML) approaches have been used for detecting AD with imaging and non-imaging data in the past years. The most widely used methods include Naïve Bayes, Support Vector Machine (SVM), and logistic regression. Naïve Bayes is a statistical method which is suitable for small dataset. SVM is a multivariate ML method which has advantage in high dimensional data classification even the data is non-linear. Logistic regression is a generating model which performs very well especially with large dataset. In this study, we chose a univariate supervised machine learning method – Naïve Bayes classifier for the following reasons: Naïve Bayes performs well even with small training data (our DIAN training dataset: 174 subjects); Compared to multivariate ML methods, it is less likely to overfit the training data with low - dimensional feature space (less than 200 features even before feature selection). Even though Naïve Bayes assumes all features used in classification which are mutually independent, it still works well for real-world data if the data doesn't meet the condition (Eisenstein and Alemi, 1993; Dyrba *et al.*, 2013).

Naïve Bayes method has been applied to predict AD progression with single and multiple modalities in recent studies (Plant *et al.*, 2010; Dyrba *et al.*, 2013; Khazaee, Ebrahimzadeh and Babajani-Feremi, 2016; Bhagya Shree and Sheshadri, 2018). Naïve Bayes is calculated based on applying Bayes' theorem and the predictive class label is determined by posterior class probability (class labels in our study: e.g. HC vs. AD, converters vs. non-converters). Posterior class probability is defined as $P(C|F_1, \dots, F_n)$ where C is class label and F is selected features. Bayesian classifiers $P(C|F_1, \dots, F_n)$ is calculated using Bayes' theorem: $P(C|F_1, \dots, F_n) = \frac{P(F_1,\dots,F_n|C)P(C)}{P(F_1,\dots,F_n)}$, where $P(C)$ is the prior probability of class, $P(F_1, \dots, F_n)$ are the probabilities of the selected features, $P(F_1, \dots, F_n|C)$ are the class conditional probabilities. In this study, all the features we used are continuous and follow Gaussian distribution (all the data were normalized described in section 2.7.1). Therefore, for each class, conditional probability was modeled with gaussian distribution: $P(f_i|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} exp\left(-\frac{(f_i-\mu_c)^2}{2\sigma_c^2}\right)$. Each subject in the test dataset was assighed to the class with higher posterior probability based on the Maximum Likelihood Estimate (MLE) with training data.

2.10. Two-stage cross validation

To test the established machine learning model (training within DIAN dataset), we did two-stage cross validation (CV): (1) At the first stage, we cross-validated with DIAN dataset (test dataset: DIAN) to select the best time period from the longitudinal disease progress for distinguishing mutation and non-mutation carriers ; (2) At the second stage, we cross-validate with ADNI dataset (test dataset: ADNI) to predict AD conversion in the early disease phase. Unlike traditional cross-validation which is usually training and testing within the same dataset, we did both intra- and inter- database cross validation to establish the machine

learning model in autosomal dominant AD and extendedly validate and apply in more common SAD.

To be more specific, we firstly trained machine learning model with DIAN dataset at different stages of ADAD. In DIAN, the disease progression of each participant can be infered according to the symptom onset their parents. Estimated year of symptom onset (EYO) is calculated by the difference between the age of participant and parental onset of symptom for both mutation and non-mutation carriers (e.g. Assuming the parent started showing symptom at 60 and the offspring is at the age of 40 at the moment, EYO is -20 based on age difference). To assess the disease stage of each participant, we defined different EYO intervals for the analysis: (1) 15 years before symptom onset $\mathrm{EYO} = (-\mathrm{Inf}, -15]$ (2) 20 to 10 years before symptom onset $\mathrm{EYO} = (-20, -10]$ (3) 15 to 5 years before symptom onset $\mathrm{EYO} = (-15, -5]$ (4) less than 10 years before symptom onset $\mathrm{EYO} = (-10, 0]$ (5) less than 5 years before and after symptom onset $\mathrm{EYO} = (-5,\ 5]$ (6) after symptom onset $\mathrm{EYO} = (0, \mathrm{Inf}]$.

For the first stage cross-validation (DIAN-DIAN), we applied repeated 10-fold CV to get stable estimation for mutation vs. Non-mutation classification(Ojala and Garriga, 2009). Repeated 10-fold CV was done through the following steps:

(1) DIAN data was randomly divided into 10 subgroups and each subgroup had same amount of data ($\mathrm{subgroup}_i$, i=1,...,10);

(2) We used 9 subgroups as training dataset and 1 subgroup as testing dataset for each fold (CV_i, i=1,...,10);

(3) In each fold (CV_i), the training dataset was randomly divided into 10 subgroups again;

(4) Step 2 was repeated only within training dataset and each fold of inner loop CV is defined as CV_{ij}(i, j=1,...,10);

(5) In the inner-loop CV_{ij}, Classification was done with different groups of features by using different IG threshold (IG percentiles from 0.5 to 0.95 with a step size of 0.05) (Dyrba *et al.*, 2015);

(6) After 10-fold inner-loop CV (each fold was done as step 5), best features (IG threshold) were selected based on AUC value and applied to the testing set of CV_i;

(7) 10 fold outer-loop CV was done for CV_1 ... CV_{10} as step 2 to step 6;

(8) We did 10 times repeated 10 fold CV (step 1 to step 7 was repeated 10 times).

After the first stage CV, established machine learning model and selected features were taken forward to the second stage validation to classify HC vs. AD and converter vs. Non-converter. For the second stage cross-validation (DIAN-ADNI), we applied repeated hold-out CV instead of 10-fold CV. Repeated hold-out CV was done through the following steps:

(1) Data and results from 1st stage CV were prepared;

(2) ADNI data was randomly divided into 10 subgroups and each subgroup had same amount of data ($subgroup_i$, i=1,...,10);

(3) We used 1 subgroup as testing dataset for each validation (CV_i, i=1,...,10);

(4) Training model of DIAN was applied to ADNI testing dataset ($subgroup_i$) for CV_i;

(5) 10 times CV was done for CV_1 ... CV_{10} as step 4;

(6) We did 10 times repeated hold-out CV (step 2 to step 5 was repeated 10 times).

For traditional 10-fold CV we used in the first stage, testing dataset in one fold will be part of the training dataset in another fold. In the second stage, we used hold-out CV because training and testing data are independent. Considering limited number of subjects in DIAN, we did repeated 10-fold CV to get stable estimation. For consistency, CV in the second stage was also repeated the same times. Two-stage CV schema is illustrated in *Figure 7*.

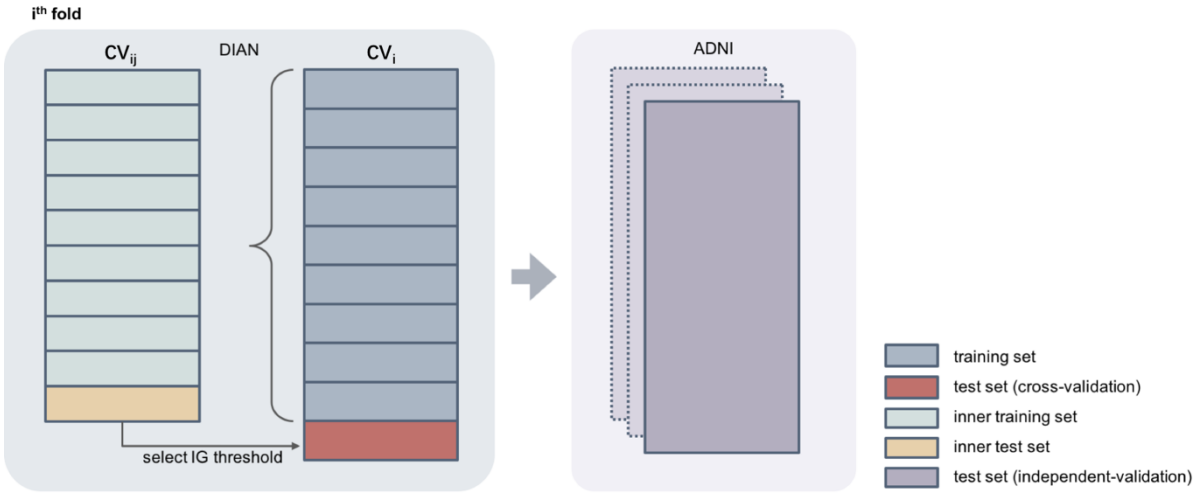

Figure 7. Cross-validation schema for feature selection and classification. Within each cross-validation fold (CV_i), another loop of cross-validation was run on the training data to estimate the best IG threshold for feature selection. The best threshold of that inner loop was the applied to the test sample of that fold (CV_i). This was run for all CV folds to determine the overall best set of features for the prediction of MC vs NC in the DIAN data. The best feature set was subsequently used as predictors tested in ADNI via Naïve Bayes, where classification result is reported as the average AUC across the repeated cross-validation.

2.11. Evaluation

To evaluate the classification result, we report the receiver operating characteristic (ROC) curve, area under the curve (AUC), sensitivity and specificity.

ROC curve and AUC are usually used to evaluate the performance of binary classifier. ROC curve is created by different thresholds and plotted with true positive rate (TPR) and false positive rate (FPR). TPR is defined as $\frac{TP}{TP+FN}$ and FPR is defined as $\frac{FP}{TN+FP}$, where TP is the number of instances correctly identified as positive class, TN is the number of instances correctly identified as negative class, FP is the number of instances incorrectly identified as positive class which belong to negative class, FN is the number of instances incorrectly identified as negative class which belong to positive class. If FPR=0 and TPR=1, it means the classifier is perfect and all the instances are correctly classified. If ROC curve is plotted as $y = x$, it means the classifier randomly classified the instances. Therefore, the closer the ROC points to the (0,1) point, the better the classifier is. AUC is the area under the ROC curve. Since the ROC curve is usually above the diagonal ($y = x$), the AUC value is between 0.5 to

1. Higher AUC value indicates better performance of classifier. ROC curve and AUC are illustrated in *Figure 8*. A big advantage of AUC is insensitivity to the data imbalance. Hence, we used AUC as a measure of classification result instead of accuracy (e.g. Assuming 90% samples belongs to the positive class and all the samples are classified as positive class, it's not a good classifier even with the accuracy of 90%). In our case, the samples in converter group are less than that in non-converter group of ADNI (HC converter/non-converter, n=18/74; MCI converter/non-converter, n=81/208). Furthermore, we also reported sensitivity and specificity as the classification result. The definition of sensitivity is the same as TPR. For the specificity, it equals to 1-FPR. All the measures (AUC, sensitivity and specificity) were averaged across multiple CV and with 95% confidence intervals reported.

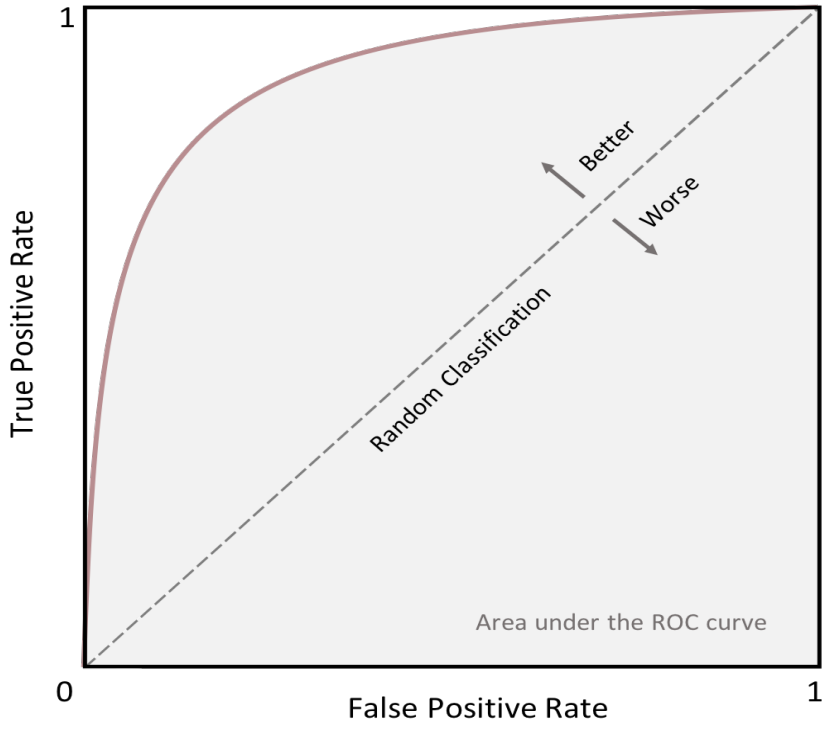

Figure 8. Example of ROC curve. The area colored in grey is area under the ROC curve.

2.12. Statistics

Demographic differences between groups were compared using two-sample t-test for continuous data and χ^2 test for categorical data (mutation vs. non-mutation carriers, HC vs. AD, HC converter vs. HC non-converter, MCI converter vs. MCI non-converter).
In the two-stage cross validation, two-sample t-tests were applied to compare the AUC between different modality combination (e.g. single modality vs. two-modality, two-modality vs. three modality and so on). In the first-stage CV, we did t-tests for the classification result of mutation vs. non-mutation carriers in DIAN and selected the best single or multiple-modality combination for the second-stage CV. In the second stage, t-tests were applied to HC vs. AD, converters vs. non-converters to compare single and multi-modality classification results. $P<0.05$ was defined as statistical significance.

2.13.Software and toolbox

Data preprocessing (normalization and standardization) were done in Python 3.6 (Spyder IDE 3.2.4). Box-cox transformation was done with SciPy library. Feature selection and classification were processed in Eclipse Java IDE (version 4.5.1) using free Weka toolbox (https://www.cs.waikato.ac.nz/ml/weka/). For statistical and other analysis, we conducted in R Statistical Computing Environment (version 1.1.414) and Matlab (version 2015b).

3 Results

3.1. Preprocessed features

We did a two-step preprocessing (normalization and standardization) for each feature of MRI, FDG-PET, neuropsychological tests and CSF biomarkers. Due to a large number of total features (N=174), we only included CSF features ($A\beta_{1\text{-}42}$, total Tau, p-tau, N=3) as examples to show the results of preprocessing.

We applied normalization as the first step of preprocessing using Box-Cox transformation to all features of all modalities. For illustrational purposes, the distribution and Quantile-Quantile plot (QQ plot) before and after normalization are shown for CSF $A\beta_{1\text{-}42}$, total Tau ,and pTau from DIAN and ADNI (*Figure 9* and *Figure 10*). For DIAN, parameters λ of Box-Cox transformation was estimated for Aβ, Tau and p-Tau respectively: $\lambda_{\mathrm{A\beta}} = 0.436,\ \lambda_{\mathrm{Tau}} = -0.122, \lambda_{\mathrm{pTau}} = -0.496$. For ADNI dataset, estimated parameter λ were: $\lambda_{\mathrm{A\beta}} = 0.437,\ \lambda_{\mathrm{Tau}} = -0.216, \lambda_{\mathrm{pTau}} = -0.281$. The results in *Figure 9* and *Figure 10* show that features become more normally distributed where points in QQ-plot become closer to the reference line after Box-Cox transformation. All CSF features are unimodally distributed except Aβ of ADNI which is from bimodal distribution. Due to technical limitations, amyloid-β_{42} CSF immunoassay was truncated to 1700 pg/mL in ADNI.

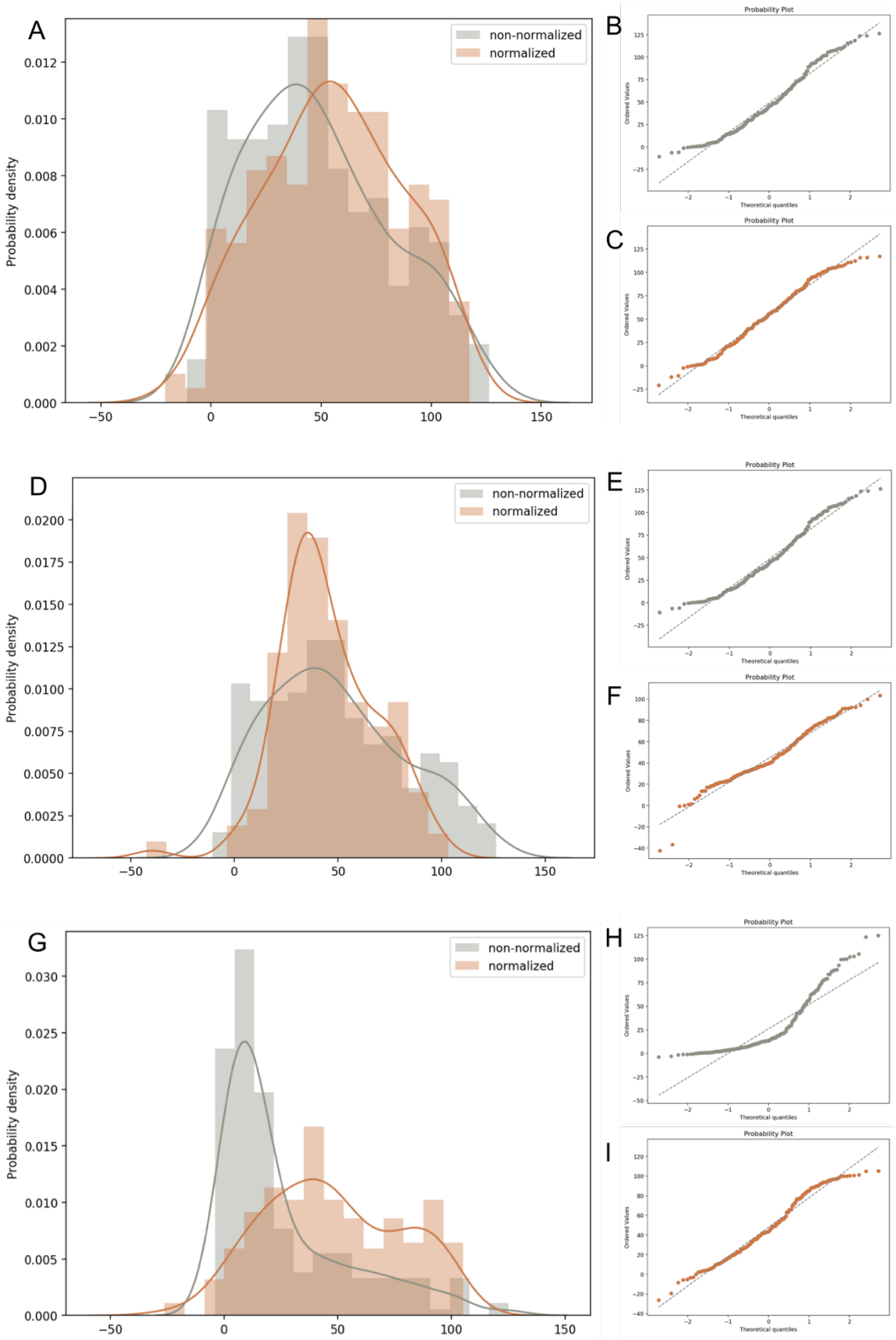

Figure 9. Histogram and normal QQ plot of DIAN CSF biomarkers. Data distribution of (A) Aβ, (D) Tau and (G) p-tau with and without normalization are shown together for comparison. QQ plots for (B, C) Aβ, (E, F) Tau and (H, I) p-tau are plotted against quantiles from a normal distribution (blue: non-normalized data, red: normalized data).

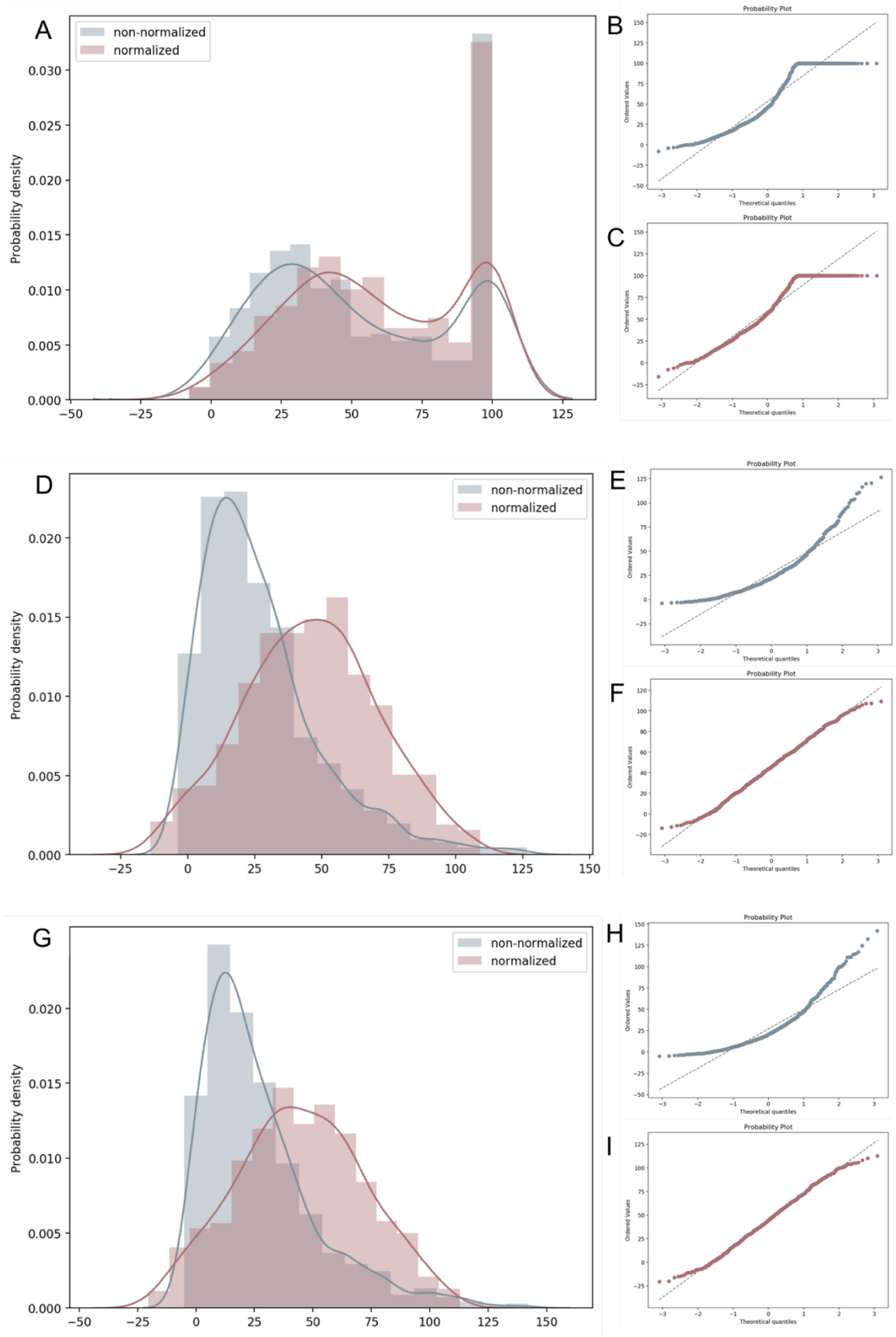

Figure 10. Histogram and normal QQ plot of ADNI CSF biomarkers. Data distribution of (A) Aβ, (D) Tau and (G) p-tau with and without normalization are shown together for comparison. QQ plots for (B, C) Aβ, (E, F) Tau and (H, I) p-tau are plotted against quantiles from a normal distribution (blue: non-normalized data, red: normalized data).

After the Box-Cox transformation, Centiloid scaling method was applied to the transformed data. *Figure 11* shows the CSF data before and after normalization and standardization for both DIAN and ADNI. Before preprocessing, CSF Aβ, Tau and pTau of DIAN and ADNI show different data ranges and all CSF data is significantly different between two datasets ($p < 0.001$) (*Figure 11A*). After preprocessing, all CSF features were converted to the similar ranges and there is no significant difference between two datasets ($p_{A\beta} = 0.060,\ p_{Tau} = 0.761, p_{pTau} = 0.150$) (*Figure 11B*).

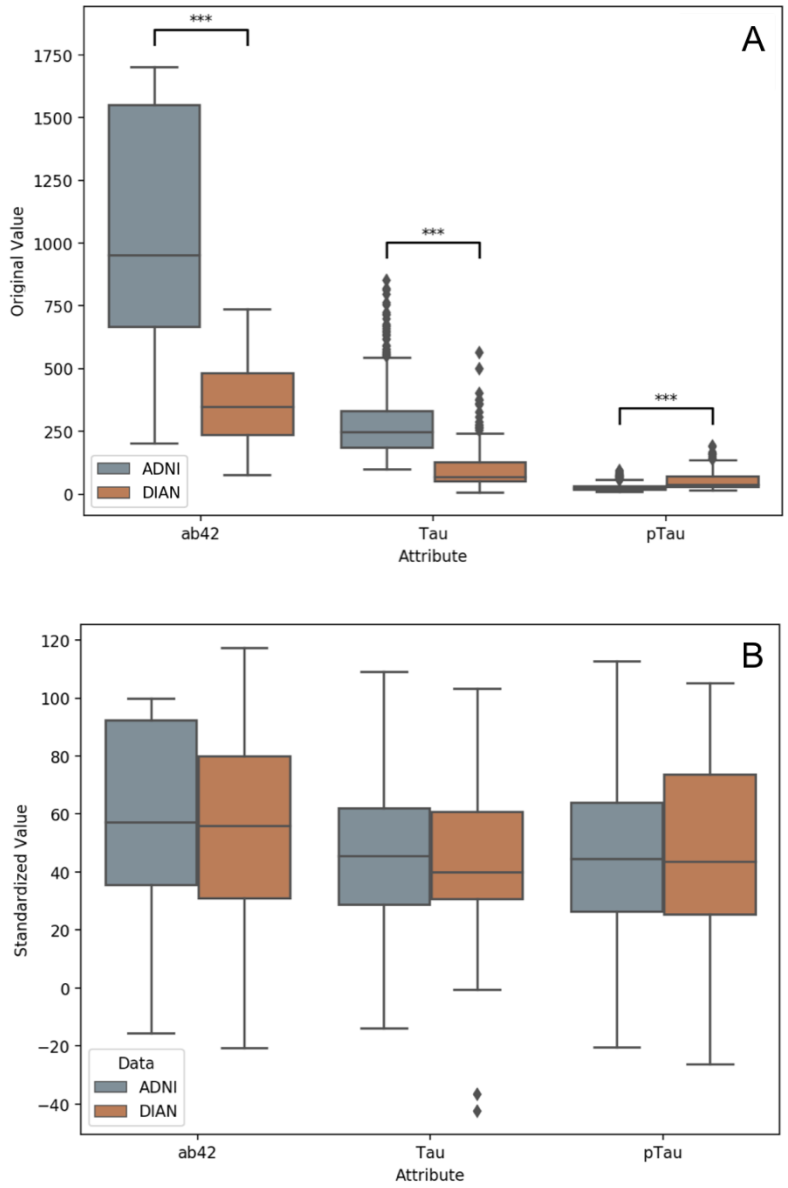

Figure 11. Box plot of CSF data for the group comparison (ADNI vs DIAN) (A) before and (B) after normalization and standardization. The top represents the 75th percentile of the data and the bottom is the 25th percentile of the data. The horizontal line in the middle is the median (50th percentile).

3.2. Establishing machine learning model within DIAN

To establish the machine learning model, we first selected the most informative features of MRI, FDG-PET, neuropsychological data and CSF biomarker respectively based on information gain method. With selected features, we applied the Naïve Bayes method to differentiate mutation and non-mutation carriers.

3.2.1 Feature selection across EYO

In DIAN, the subjects were categorized in different 10-year EYO intervals. We did feature selection and classification for each EYO interval sequentially. For the cross-validation, we did repeated (10 times) 10-fold CV which has been described in details in the method part. The selected features whose IG values were above the thresholds in more than half CV folds (larger than the IG threshold in more than 10*10/2 folds) were shown in *Figure 12* to *Figure 15*.

Using MRI, selected features for differentiating mutation carriers and non-carriers at more than 15 years before the onset of symptoms were mainly in the central of left hemisphere: left paracentral lobule and postcentral gyrus (*Figure 12A*). By $\mathrm{EYO} = (-20, -10]$, selected features of grey matter volumes were left putamen and right lingual (*Figure 12B*). ROIs from the central regions in both medial and lateral sides (left paracentral lobule and postcentral gyrus, right supramarginal gyrus) were selected at the $\mathrm{EYO} = (-15, -5]$ (*Figure 12C*). At the EYO interval $(-10, 0]$, right accumbens and precuneus were selected as the most informative features (*Figure 12D*). By $\mathrm{EYO} = (-5, 5]$ which is around the year of symptom onset, more features were selected including left putamen, left and right hippocampus, right banks of the superior temporal sulcus, right cuneus and precuneus (*Figure 12E*). At the years after symptom onset, similar features as $(-5, 5]$ were selected: left and right hippocampus, left amygdala, left and right postcentral gyrus, right cuneus (*Figure 12F*). From the years

approaching symptom onset to years after the onset, selected features of GM volumes for classification were mostly memory-related such as hippocampus and precuneus.

Using FDG-PET, selected features at more than 15 years before the onset of symptoms were left caudal anterior cingulate cortex, right fusiform and superior parietal cortex (*Figure 13A*). By $\mathrm{EYO} = (-20, -10]$, selected features of FDG-PET were left and right hippocampus, left amygdala and right frontal pole (*Figure 13B*). Three ROIs from the right hemisphere (pallidum, hippocampus, and accumbens) were selected at the $\mathrm{EYO} = (-15, -5]$ (*Figure 13C*). At the EYO interval $(-10, 0]$, left inferior parietal, right cuneus, and isthmus-cingulate cortex were selected as the most informative features (*Figure 13D*). By $\mathrm{EYO} = (-5, 5]$ which is around the year of symptom onset, features were selected including left and right precuneus, right isthmus-cingulate cortex (*Figure 13E*). In the years after symptom onset, more features were selected: left inferior and superior parietal cortex, left and right isthmus-cingulate cortex, right precuneus (*Figure 13F*). From the years approaching symptom onset to years after the onset, selected features of FDG-PET for classification were mostly memory-related such as parietal and cingulate cortex.

Using the data of neuropsychological tests, selected features at more than 15 years before the onset of symptoms were mainly related to semantic and episodic memory: Boston Naming Test, Animal Fluency Test and Word List Delayed Recall. For $\mathrm{EYO} = (-20, -10]$, selected features of neuropsychological data were Boston Naming Test, Animal Fluency Test and Word List Immediate Recall. Features related to executive function (Animal Fluency Test, Trail Making Test (A&B)) were selected at the $\mathrm{EYO} = (-15, -5]$. At the EYO interval $(-10, 0]$, Boston Naming Test, Animal Fluency Test, and MMSE were selected as the most informative features. By $\mathrm{EYO} = (-5, 5]$ which is around the year of symptom onset, features were selected including Boston Naming Test, Trail Making Test (A), Word List Delayed, and Immediate Recall. At the years after symptom onset, three features were selected: Boston

Naming Test and Trail Making Test (A&B). Neuropsychological features of different EYO intervals are shown in *Figure 14*.

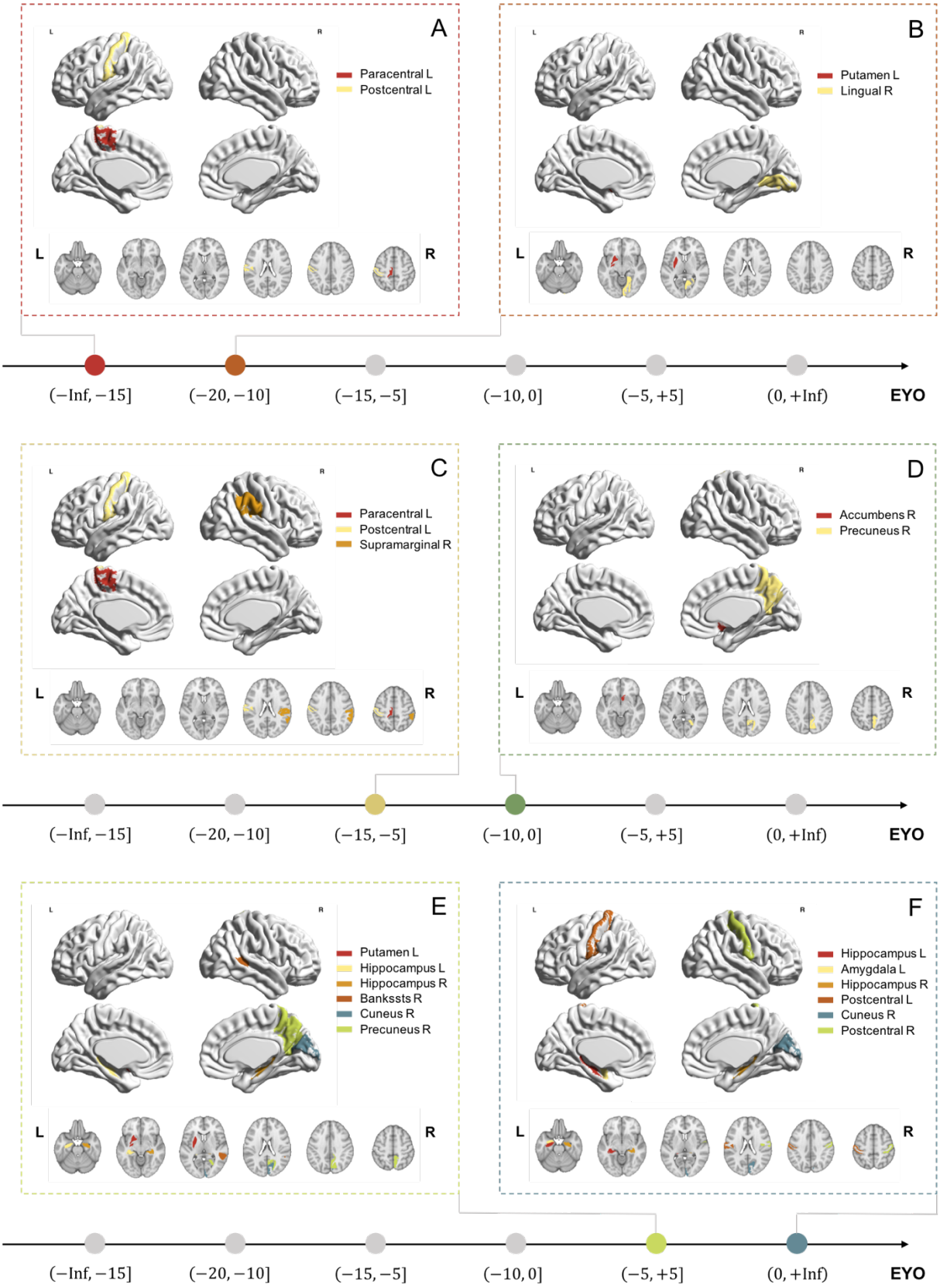

Figure 12. Desikan-Killiany atlas of selected MRI features at different EYO intervals: (A) $(-Inf, -15]$ *(B)* $(-20, -10]$ *(C)* $(-15, -5]$ *(D)* $(-10, 0]$ *(E)* $(-5,\ 5]$ *(F)* $(0, Inf)$.

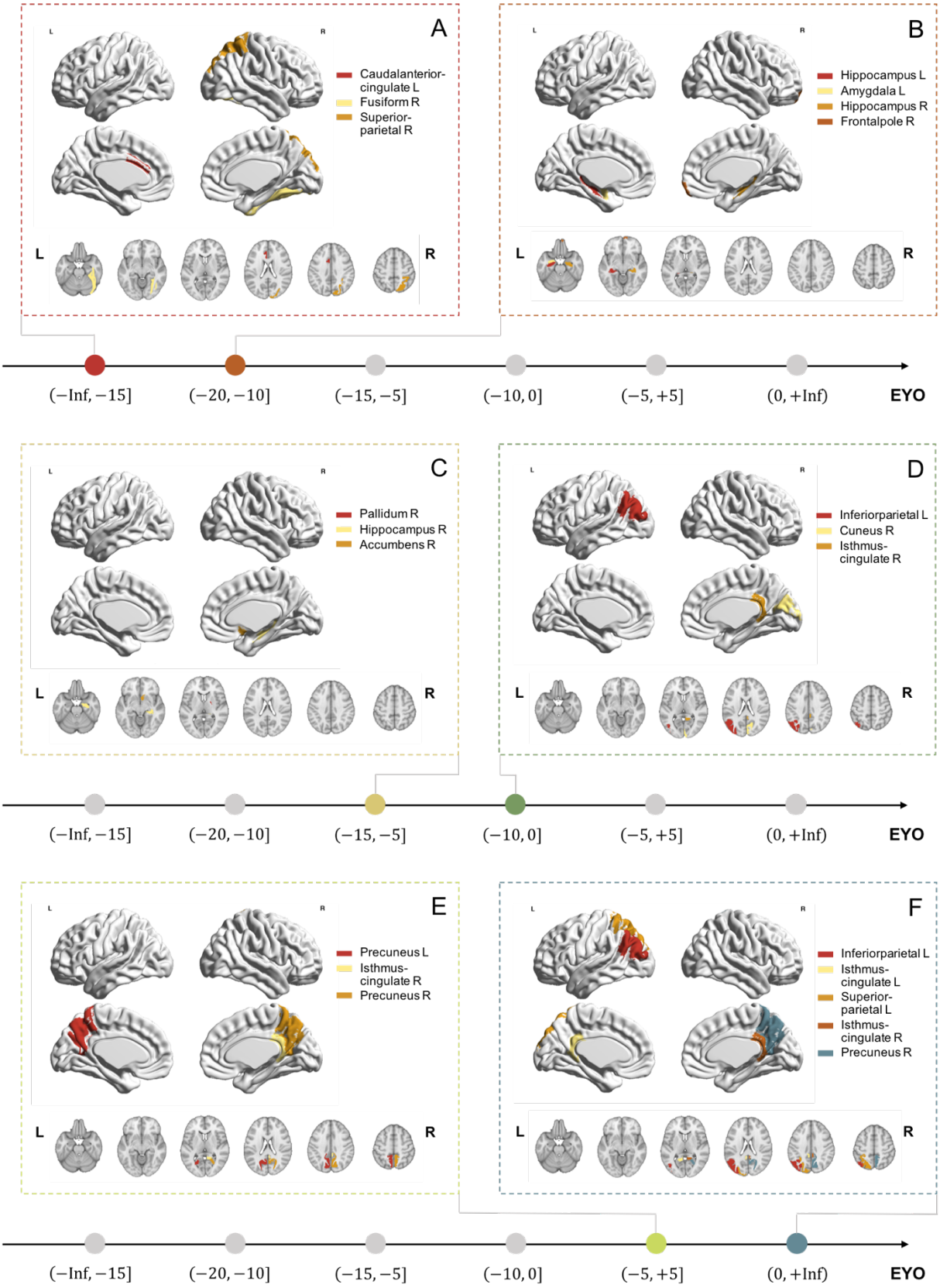

Figure 13. Desikan-Killiany atlas of selected FDG-PET features at different EYO intervals: (A) $(-Inf, -15]$ *(B)* $(-20, -10]$ *(C)* $(-15, -5]$ *(D)* $(-10, 0]$ *(E)* $(-5,\ 5]$ *(F)* $(0, Inf)$.

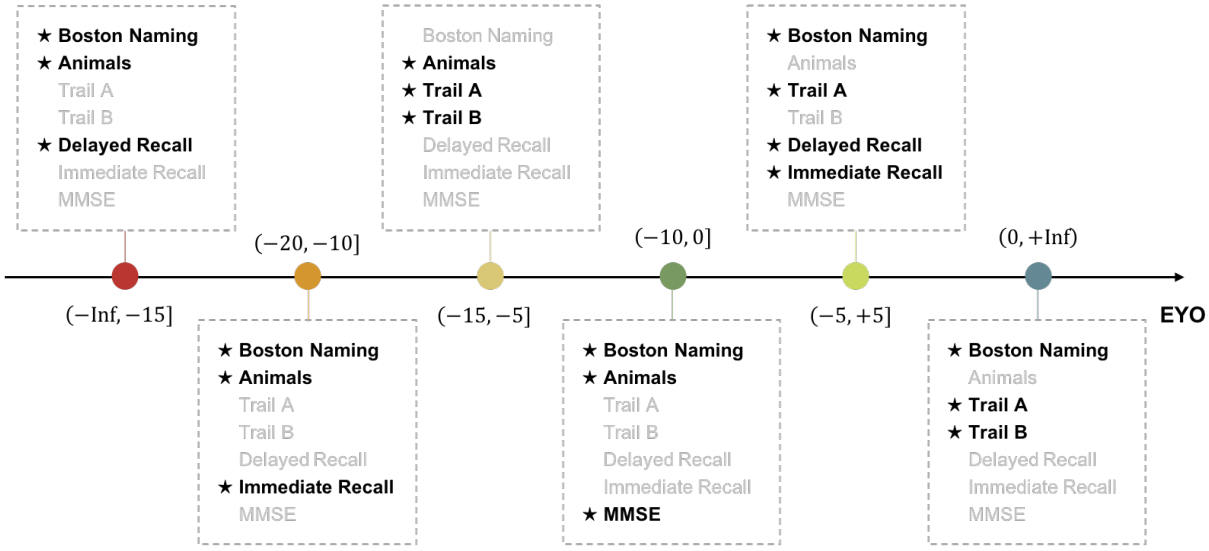

Figure 14. Selected neuropsycholocal features acrosst different EYO intervals

Using CSF biomarkers, selected feature at more than 15 years before the onset of symptoms was pTau. By EYO $= (-20, -10]$, the selected feature of CSF biomarkers was pTau which is the same as earlier EYO interval. Aβ and pTau were selected at the EYO $= (-15, -5]$. At the EYO interval $(-10, 0]$, Aβ and pTau were selected as the most informative features. By EYO $= (-5, 5]$ which is around the year of symptom onset, the selected feature was pTau. At the years after symptom onset, two features were selected: Aβ and pTau. From the very early years before symptom onset, pTau was shown as an informative feature for differentiating mutation carriers and non-carriers until the years after symptom onset which indicates pTau is a prominent biomarker for the early diagnosis of AD. CSF features of different EYO intervals are shown in *Figure 15*.

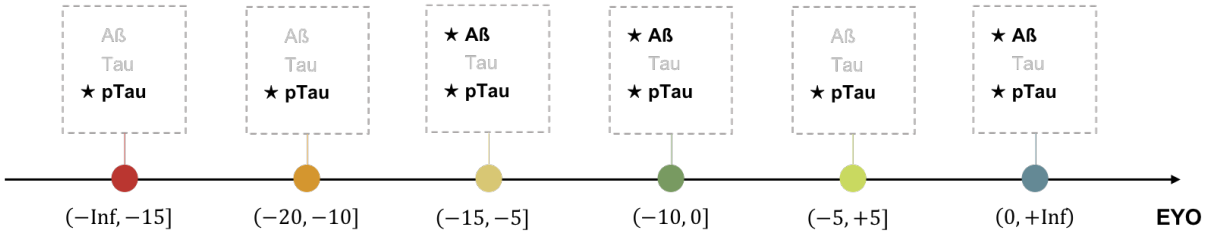

Figure 15. Selected CSF features at different EYO intervals

3.2.2 Classification of ADAD for different EYO intervals

With selected features described in 3.2.1, the classification was done using Naïve Bayes method for the separation between mutation and non-mutation carriers with single and multiple modalities across different EYO intervals. The results of the cross-validation are shown in *Figure 16*, *Table 9* and *Table 10*. In general, classification results within DIAN became better when the time gets closer to the symptom onset.

The classification results (AUC) of single modality are shown in *Figure 16A*. For MRI, AUC generally increased across EYO intervals (except the EYO interval $(-10, 0]$ which was a bit decreased comparing to $(-15, -5]$). AUC became larger than 85% (AUC = 90.00%) at the EYO interval [0, +Inf). For FDG-PET, AUC increased across EYO intervals from the EYO interval of $(-20, -10]$. From the EYO interval of $(-5, 5]$, the AUCs of FDG-PET were better than the AUCs of MRI and AUC achieved higher than 85% (AUC = 89.50, 92.25%). For neuropsychological data, AUC increased across EYO intervals across all the EYO intervals and approached 85% at the EYO of $(-10, 0]$ (AUC = 84.11%). From the EYO interval of $(-10, 0]$, the AUCs of psychological data were better than the AUCs of MRI. For CSF, AUC increased across EYO intervals from the EYO interval of $(-15, -5]$ and reached 89.17% at $(-10, 0]$ (>85%). The AUCs of CSF biomarkers were larger than the other single modality across most time intervals (except the EYO interval $(-15, -5]$ which was lower than MRI). Across all EYO intervals, highest AUC was observed for CSF (97.75% at $(0, +Inf)$).

The classification results (AUC) of two modality combinations are shown in *Figure 16B*. For MRI-FDG (M-F), AUCs were higher than both MRI and FDG-PET across most of the EYO intervals except $(-15, -5]$. For MRI-PSY (M-P), AUCs were higher than both MRI and neuropsychology across most of the EYO intervals except $(-20, -10]$. For MRI-CSF (M-C), AUCs were higher than both MRI and CSF across most of the EYO intervals except $(-5, 5]$.

The AUC of M-C became clinical relevent from $(-20, -10]$ (AUC = 91.67%) which was earlier than every single modality. For FDG-PSY (F-P), AUCs were higher than both FDG-PET and neuropsychology across most of the EYO intervals except $(0, +Inf)$. For FDG-CSF (F-C), AUCs were higher than both FDG-PET and CSF across most of the EYO intervals except $(-10, 0]$ and $(-5, 5]$ where CSF had higher AUC values. For PSY-CSF (P-C), AUCs were not higher than CSF at $(-Inf, -15]$, $(-20, -10]$, $(-15, -5]$ and $(-5, 5]$.

The classification results (AUC) of three and four (all) modality combinations are shown in *Figure 16C*. For MRI-FDG-PSY (M-F-P), AUCs were higher than every single one of these three modalities across all of EYO intervals but not higher than all two modality combinations from these three modalities. For MRI-FDG-CSF (M-F-C), AUCs were higher than almost every single one of these three modalities across all of EYO intervals except CSF of $(-5, 5]$ and not higher than all two modality combinations from these three modalities. For MRI-PSY-CSF (M-P-C), AUCs were higher than every single one of these three modalities across all of EYO intervals and equal or higher than most two modality combinations from these three modalities (except M-C of $(-20, -10]$). For FDG-PSY-CSF (F-P-C), AUCs were higher than most single ones across all of EYO intervals except CSF of $(-20, -10]$ and $(-15, -5]$, but not higher than all two modality combinations from these three modalities. Combining all four modalities, the results showed that AUCs of all-modality-combination were higher than every single modality across all of EYO intervals. However, compared to two and three modality combinations, all modality combinations did not improve classification results for all EYO intervals.

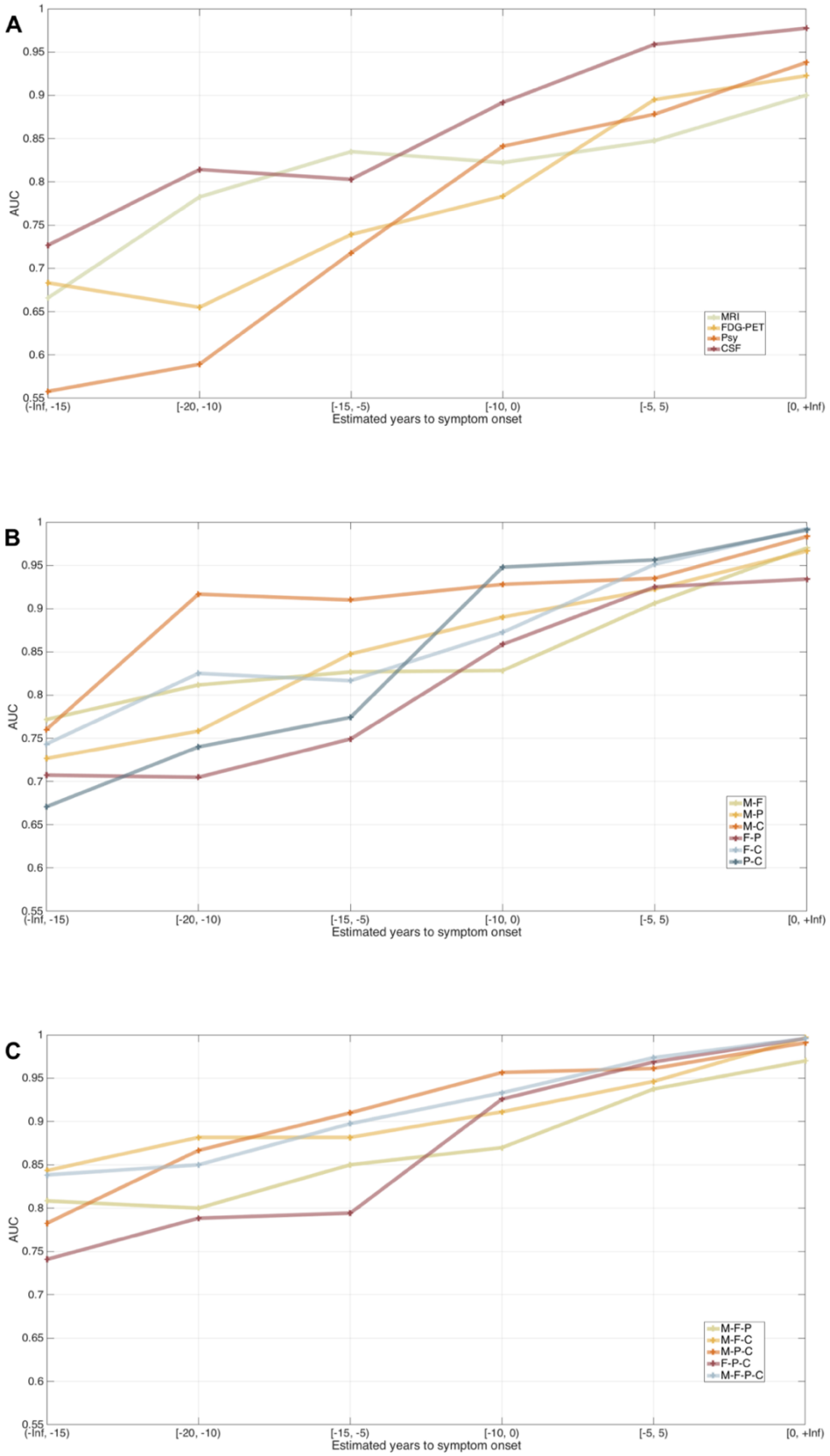

Figure 16. AUC for the discrimination of mutation carriers vs. non-mutation carriers of (A) single modality (B) two-modality combination (C) three-modality and all combination across different EYO intervals.

Table 9: AUC(95%CI) for the discrimination of mutation carriers vs. non-mutation carriers of single modality, two-modality combination, three-modality and all combination across different EYO intervals

Modality/ Modalities	Classification results (AUC) of different EYO intervals					
	(-Inf, -15)	[-20, -10)	[-15, -5)	[-10, 0)	[-5, 5)	[0, +Inf)
MRI	66.58±2.60	78.25±2.50	83.50±2.26	82.22±1.71	84.75±1.80	90.00±1.67
FDG	68.33±2.88	65.50±2.86	73.92±2.66	78.32±2.15	89.50±1.50	92.25±1.33
PSY	55.79±3.11	58.92±2.68	71.79±2.63	84.11±1.93	87.81±1.62	93.79±1.20
CSF	72.67±2.42	81.42±2.07	80.29±2.52	89.17±1.25	95.88±0.98	97.75±0.76
M-F	77.17±2.34	81.17±2.34	82.67±2.14	82.83±1.99	90.63±1.30	97.00±0.86
M-P	72.67±2.60	75.83±2.31	84.75±2.08	89.01±1.52	92.25±1.49	96.67±0.91
M-C	76.00±2.48	91.67±1.22	91.00±1.66	92.81±1.19	93.50±1.16	98.33±0.69
F-P	70.75±2.76	70.50±2.51	74.92±2.64	85.86±1.84	92.50±1.27	93.42±1.16
F-C	74.33±2.68	82.50±1.90	81.67±2.44	87.25±1.54	95.13±0.87	99.25±0.38
P-C	67.08±2.70	74.00±2.16	77.42±2.61	94.79±0.89	95.63±0.89	99.08±0.47
M-F-P	80.83±2.25	80.00±2.15	85.00±2.09	86.99±1.77	93.75±1.13	97.00±0.89
M-F-C	84.33±1.95	88.17±1.63	88.17±1.90	91.11±1.33	94.63±0.91	99.67±0.23
M-P-C	78.25±2.54	86.67±1.61	91.00±1.80	95.65±0.96	96.13±0.81	99.08±0.47
F-P-C	74.08±2.60	78.83±1.95	79.42±2.58	92.58±1.22	96.88±0.67	99.58±0.30
M-F-P-C	83.83±2.05	85.00±1.76	89.75±1.82	93.31±1.23	97.38±0.62	99.58±0.30

For each modality or modality combination, the average AUC with SE of repeated cross-validation is presented. Abbreviations: M-F = MRI & FDG; M-P = MRI & PSY; M-C = MRI & CSF; F-P = FDG & PSY; F-C = FDG & CSF; P-C = PSY & CSF; M-F-P = MRI & FDG &PSY; M-F-C = MRI & FDG & CSF; M-P-C = MRI & PSY & CSF; F-P-C = FDG &PSY & CSF; M-F-P-C = MRI & FDG &PSY & CSF.

Table 9: Sensitivity and specificity for the discrimination of mutation carriers vs. non-mutation carriers of single modality, two-modality combination, three-modality and all combination across different EYO intervals

Modality/ Modalities	Classification results (sensitivity, specificity) of different EYO intervals					
	(-Inf, -15)	[-20, -10)	[-15, -5)	[-10, 0)	[-5, 5)	[0, +Inf)
MRI	85.33, 60.00	83.33, 75.17	93.33, 74,17	90.00, 68.33	90.25, 85.50	91.75, 93.50
FDG	88.33, 62,50	88.67, 59.67	93.33, 64.33	90.00, 59.17	92.00, 90.50	95.75, 82.00
PSY	85.33, 47.00	87.00, 48.17	93.33, 58.83	90.00, 70.17	90.00, 90.50	96.25, 82.50
CSF	88.33, 62.50	91.33, 73.33	94.67, 74.00	90.00, 82.50	95.50, 95.00	96.83, 98.50
M-F	83.33, 72.50	83.33, 75.67	93.33, 72.67	90.00, 68.83	92.00, 91.50	95.75, 93.00
M-P	83.33, 66.00	83.33, 71.50	93.33, 77.33	92.08, 91.17	93.50, 94.50	95.83, 99.00
M-C	83.33, 72.00	94.67, 89.83	98.00, 86.50	92.25, 96.00	95.50, 94.50	98.33, 98.50
F-P	83.33, 63.50	83.33, 63.00	93.33, 65.00	90.50, 86.83	92.50, 96.00	96.75, 84.50
F-C	83.33, 68.50	83.33, 76.33	93.33, 75.50	90.17, 90.17	94.25, 96.00	98.75, 100
P-C	83.33, 57.50	83.67, 67.17	93.33, 69.50	92.67, 98.67	95.25, 98.00	99.08, 99.50
M-F-P	83.33, 77.00	83.33, 74.33	93.33, 76.17	90.00, 75.00	93.75, 96.00	96.75, 93.00
M-F-C	83.33, 81.50	83.33, 85.83	93.33, 82.17	90.25, 82.50	94.50, 96.00	96.00, 99.00
M-P-C	83.33, 75.00	83.33, 83.50	93.33, 88.33	90.00, 90.67	94.75, 99.00	96.75, 98.50
F-P-C	83.33, 68.00	83.33, 73.83	93.33, 72.17	90.00, 84.83	96.00, 98.00	97.00, 99.50
M-F-P-C	83.33, 81.00	83.33, 81.17	93.33, 85.00	90.25, 85.83	96.75, 98.00	97.00, 99.50

For each modality or modality combination, the average sensitivity and specificity of repeated cross-validation is presented. Abbreviations: M-F = MRI & FDG; M-P = MRI & PSY; M-C = MRI & CSF; F-P = FDG & PSY; F-C = FDG & CSF; P-C = PSY & CSF; M-F-P = MRI & FDG &PSY; M-F-C = MRI & FDG & CSF; M-P-C = MRI & PSY & CSF; F-P-C = FDG &PSY & CSF; M-F-P-C = MRI & FDG &PSY & CSF.

3.2.3 Optimal classification model of DIAN

Before further cross-validation in ADNI, we selected the prediction model from classification models of different EYO intervals. Based on the classification results and sample size, we selected the machine learning model trained for EYO interval $(-5, +Inf)$ because 1) the classification result for mutation vs. non-mutation carriers was within the clinical relevant accuracy (AUC > 85%) and 2) the classification model was established to predict the early stage of AD and 3) sufficient sample size compared to one EYO interval.

For the EYO interval of $(-5, +Inf)$, we also did feature selection using information gain (IG) method. For MRI, the selected features included left and right hippocampus, left putamen and amygdala, right precuneus, right postcentral gyrus, and cuneus. The average IG value of all ROI features (41 ROIs for each hemisphere) and MRI features after selection are shown in *Figure 17*.

For FDG-PET, the selected ROI features were predominantly located within posterior parietal cortex of the default mode network including left inferior and superior parietal cortex, left and right isthmus-cingulate cortex, left and right precuneus. The average IG value of all ROI features and selected FDG-PET features are shown in *Figure 18*.

For neuropsychological tests, the most discriminative features were Boston Naming Test, Trail Making Test (A&B) and Word List Immediate Recall. For CSF biomarkers, Aβ and pTau were selected as the most informative features. The average IG value of all neuropsychological and CSF features before and after selection are shown in *Figure 19*.

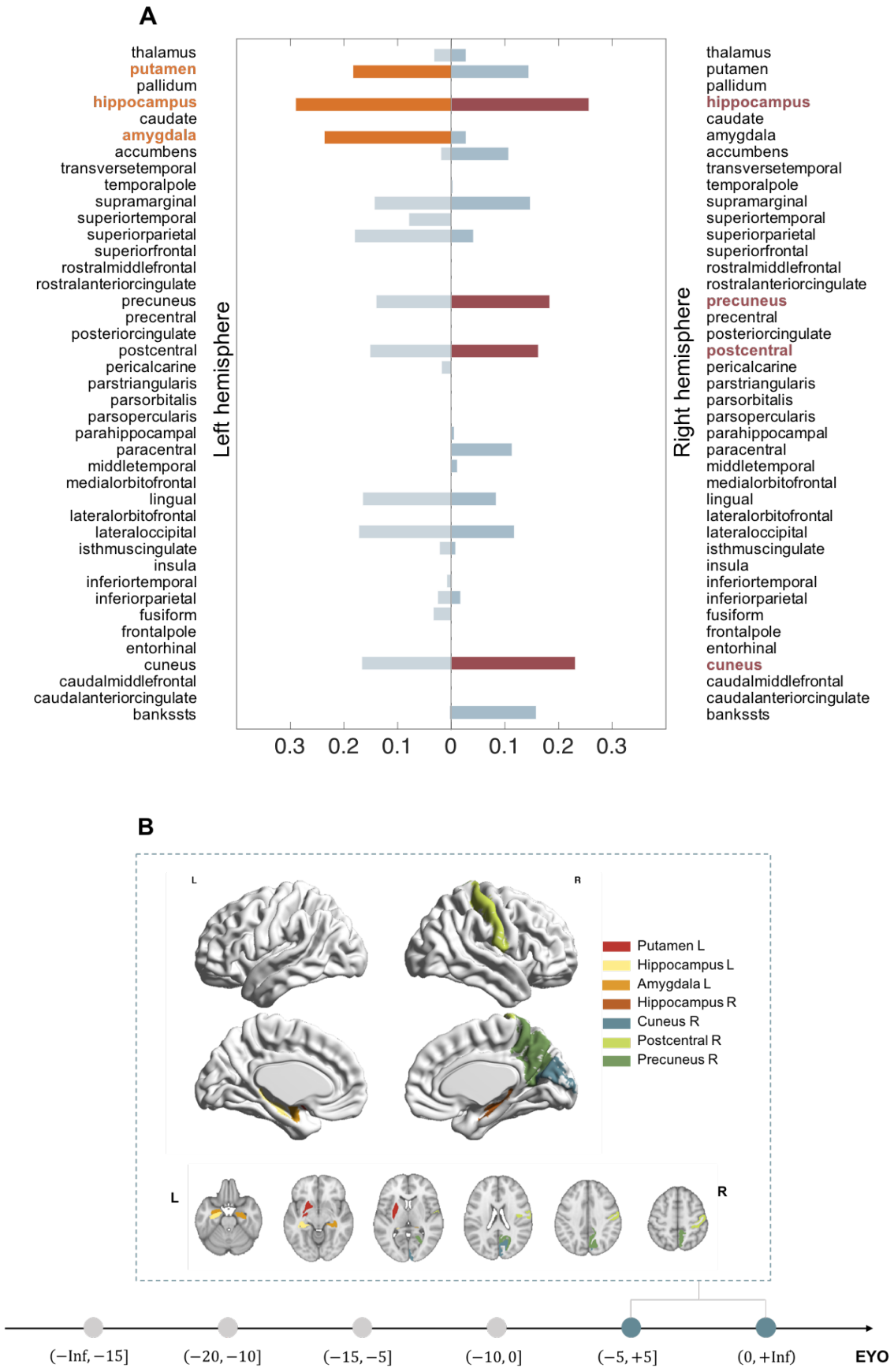

Figure 17. (A) Bar graph of mean IG value for each MRI feature (selected features were colored with orange for the left hemisphere and red for the right hemisphere). (B)Selected MRI features from Desikan-Killiany atlas at the EYO interval of $(-5, +Inf)$.

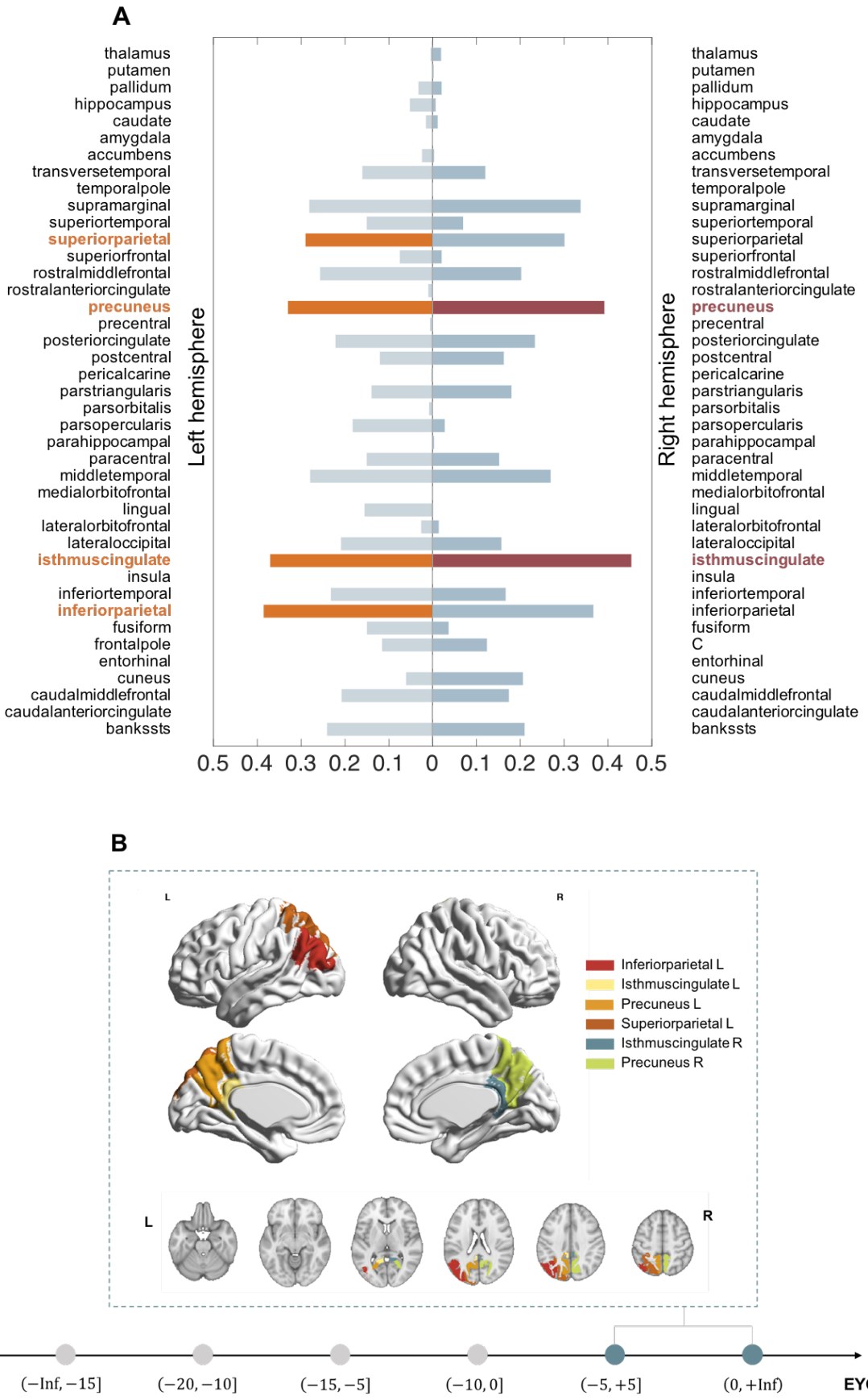

Figure 18. (A) Bar graph of mean IG value for each FDG-PET feature (selected features were colored with orange for the left hemisphere and red for the right hemisphere). (B)Selected FDG-PET features from Desikan-Killiany atlas at the EYO interval of $(-5, +Inf)$.

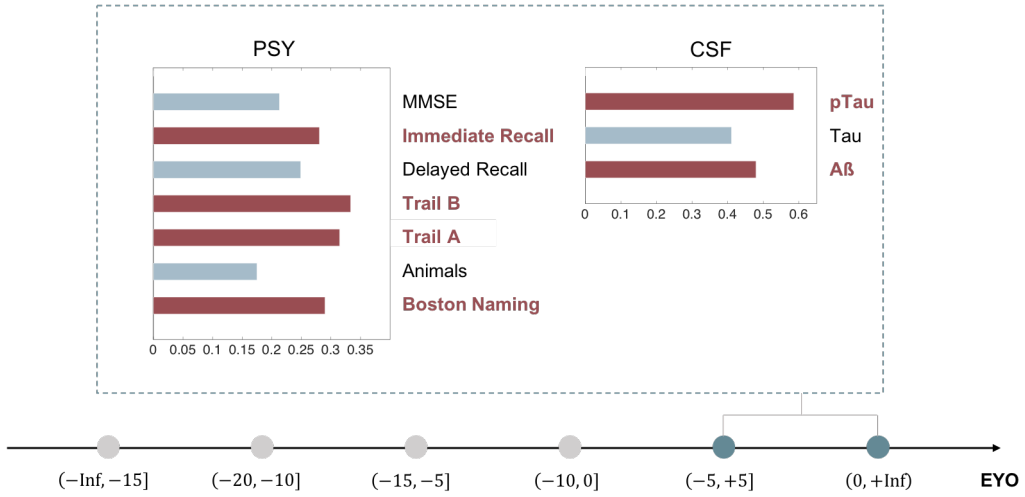

Figure 19. Bar graph of mean IG value for each PSY and CSF feature (selected features were colored with red) at the EYO interval of $(-5, +Inf)$.

Classification results of single and multiple modalities for distinguishing mutation and non-mutation carriers at the EYO interval of $(-5, +Inf)$ are shown in *Figure 20* and *Table 10*. For the single modality, the best classification result was using CSF biomarkers with the AUC of 93.75%. Concatenating features from two modalities, the best performance was combining neuropsychological data and CSF biomarkers which was significantly better than every single modality. Besides PSY-CSF, the AUC of FDG-CSF was also significantly improved compared to every single modality. For the rest two-modality combinations (MRI-FDG, MRI-PSY, FDG-PSY, MRI-CSF), classification results were significantly better than the single modality of MRI and FDG-PET, but not better than PSY or CSF. The best combination among three-modality combinations was FDG-PSY-CSF, which was significantly better than the all two-modality combinations except PSY-CSF. Other than that, the AUC of MRI-PSY-CSF was significantly higher than MRI-FDG, MRI-PSY, FDG-PSY, and MRI-CSF. MRI-FDG-PSY performed significantly better than MRI-FDG, MRI-PSY, and FDG-PSY. However, the result of three-modality MRI-FDG-CSF didn't improve significantly from any two-modality combination. Combining all four modalities, the AUC was significantly better than all three-modality combinations except FDG-PSY-CSF.

Table 10: Mean AUC, sensitivity and specificity for the discrimination of mutation carriers vs. non-mutation carriers of single modality, two-modality combination, three-modality and all combination at the EYO interval of $(-5, +Inf)$.

Modality/ Modalities	Classification results of EYO interval $(-5, +Inf)$	
	Mean AUC (95%CI)	Sensitivity, Specificity
MRI	86.38 (84.06, 88.70)	85.97, 80.50
FDG	90.07 (88.02, 92.12)	88.50, 88.08
PSY	93.61 (92.05, 95.17)	91.03, 91.58
CSF	93.75 (91.91, 95.59)	92.53, 95.00
M-F	92.54 (90.71, 94.37)	89.80, 92.92
M-P	94.19 (92.54, 95.84)	91.33, 93.42
M-C	95.24 (93.94, 96.54)	92.53, 97.08
F-P	94.87 (93.33, 96.41)	92.40, 95.25
F-C	95.97 (94.62, 97.32)	93.53, 97.08
P-C	96.92 (95.67, 98.17)	95.47, 98.33
M-F-P	96.25 (94.83, 97.67)	94.37, 95.17
M-F-C	95.93 (94.62, 97.24)	93.23, 97.67
M-P-C	97.03 (95.96, 98.10)	94.50, 98.75
F-P-C	98.08 (97.23, 98.93)	96.07, 98.75
M-F-P-C	98.18 (97.29, 99.07)	96.67, 99.67

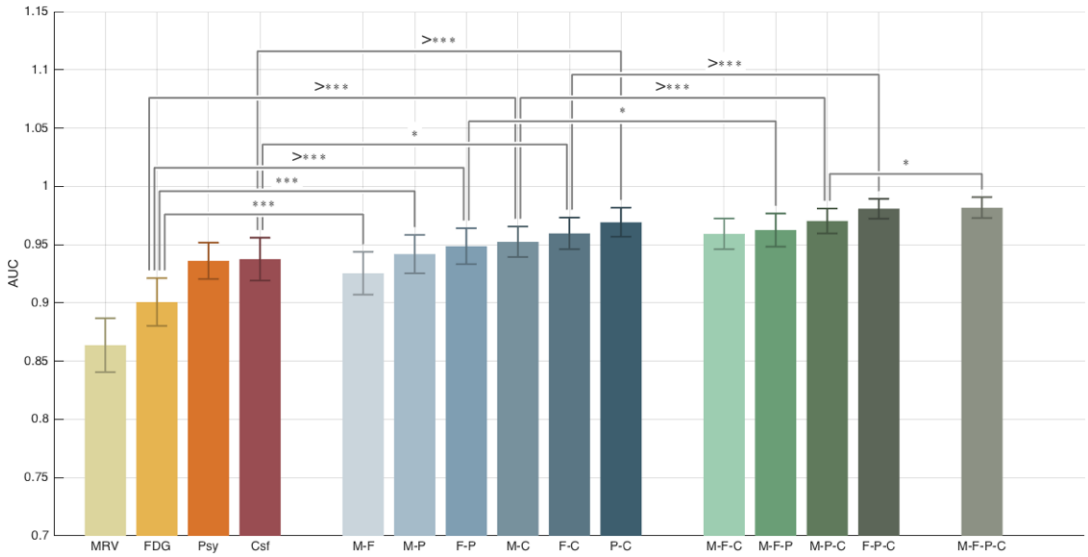

Figure 20. AUC of single and multiple modalities (classification of mutation carriers and non-mutation carriers). The significant increasing of AUC is only marked for the highest one (e.g. The AUC of combining psychological data with CSF biomarkers is significantly higher than any single modality and it is only marked with the modality with highest AUC: CSF biomarkers).

3.3. Cross-validation in ADNI

With the machine learning model trained in DIAN, we further applied to the ADNI dataset to validate in the more common SAD. Classification models of every single modality (MRI, FDG, PSY, CSF) and the multi-modality combinations with best results were used in ADNI dataset.

3.3.1 Discrimination between HC and AD in ADNI

We used three single modality models (MRI, FDG, CSF), the two-modality combination (FDG-CSF) and the three-modality combination (MRI-FDG-CSF) models with the best performance in DIAN to further apply to the ADNI dataset. Note that neuropsychological data was not used for discriminating HC versus AD since it provided diagnostic information which may lead to circularity. The ROC curves and classification results (AUC, sensitivity and specificity) of single- and multi- modality are shown in *Figure 21* and *Table 11*. Among the single modality models, the FDG-PET model had the best result (AUC = 91.96%). The two-modality combination FDG-CSF (best two-modality model in DIAN), achieved AUC of 94.80%. The three-modality combination MRI-FDG-CSF showed AUC of 96.00%.
In comparison with each single modality model, FDG-CSF performed significantly better. The three-modality MRI-FDG-CSF even had significantly higher AUC than FDG-CSF model. In general, MRI-FDG-CSF showed the best result among all single- and multi-modality models.

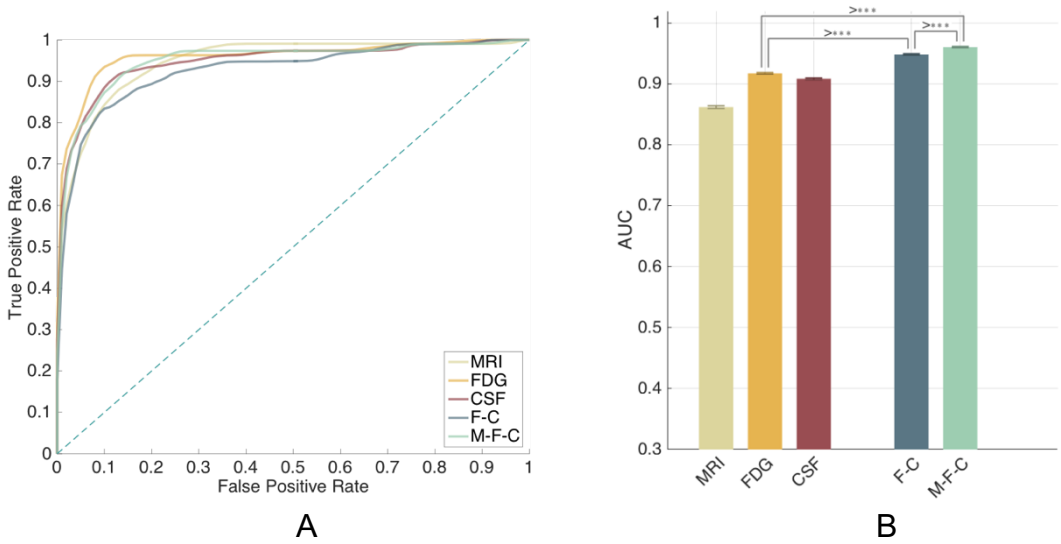

Figure 21. Classification result for HC versus AD in SAD based on ADAD Naïve Bayes model with MR volume, FDG-PET and CSF biomarkers and the multi-modalities combination with best performance in ADAD. (a) ROC curves of classification: HC vs. AD. (b) AUC of classification: HC vs. AD. The significant increasing of AUC is only marked for the highest one. AUC of F-C and M-F-C is significant lager than FDG-PET and it is also significant larger than MRI and CSF (AUC: FDG > CSF > MRI)

Table 11: *Classification result (AUC, sensitivity and specificity) for HC versus AD of ADNI data with single modality and corresponding best multi-modality combination tested in DIAN.*

Modality/Modalities	Mean AUC (95%CI)	Sensitivity, Specificity
MRI	86.17 (85.94, 86.40)	83.11, 77.81
FDG	91.69 (91.55, 91.82)	83.38, 88.12
CSF	90.81 (90.66, 90.96)	84.00, 87.91
F-C	94.80 (94.69, 94.90)	89,11, 90.62
M-F-C	96.00 (95.89, 96.11)	93.35, 91.40

3.3.2 Classification of HC converters and non-converters in ADNI

We used all single modality models (MRI, FDG, PSY, CSF), the two-modality combination (PSY-CSF) and the three-modality combination (FDG-PSY-CSF) models with the best performance in DIAN, and all-modality combination model (MRI-FDG-PSY-CSF) to distinguish HC converters and non-converters of ADNI dataset. The ROC curves and classification results (AUC, sensitivity, and specificity) of single- and multi- modality are shown in *Figure 22* and *Table 12*. Among the single modality models, the neuropsychological

model had the best result (AUC = 68.17%), and CSF model (best single model in DIAN) showed the AUC of 49.57% which was not better than the other single modalities. The two-modality combination PSY-CSF (best two-modality model in DIAN), achieved AUC of 58.22%. The three-modality combination FDG-PSY-CSF (best three-modality model established in DIAN), showed AUC of 63.28%. All-modality combination (MRI-FDG-PSY-CSF) achieved AUC of 62.69%.

Compared with each single modality model, FDG-CSF performed only significantly better than CSF model. The three-modality FDG-PSY-CSF showed significantly higher AUC than PSY-CSF model as well as FDG and CSF single-modality models. Four-modality-combination didn't show better performance compared to three- modality model.

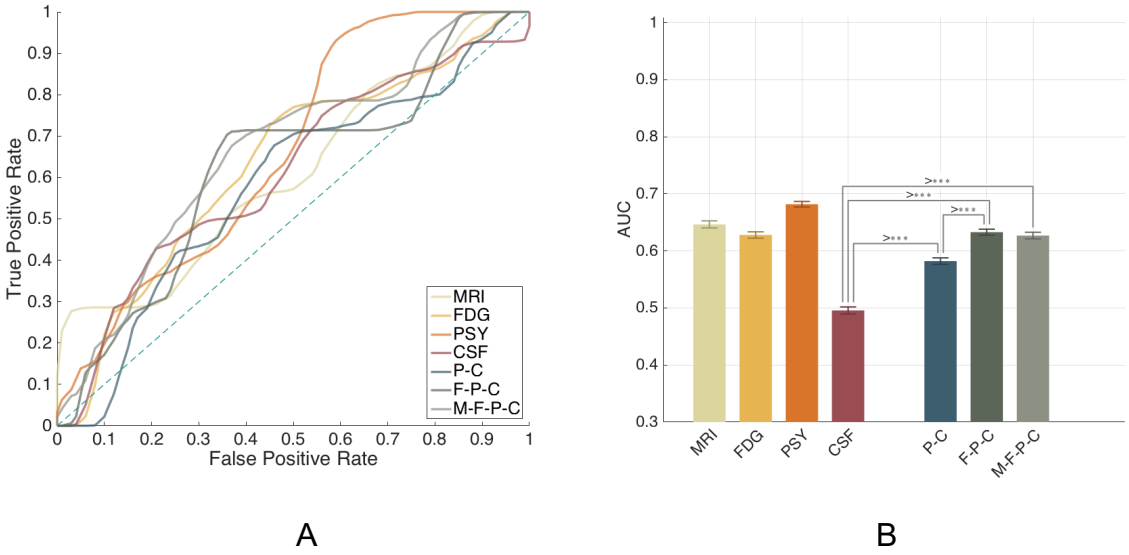

Figure 22. Classification result for HC Converters (HC-C) versus HC non-Converters (HC-NC) in SAD based on ADAD Naïve Bayes model with all four modalities: MR volume, FDG-PET, Psychological data and CSF biomarkers and the multi-modality combinations with best performance in ADAD. (a) ROC curves of classification: HC-C vs. HC-NC. (b) AUC of classification: HC-C vs. HC-NC. The significant increasing of AUC is only marked for the highest one. (AUC: PSY > MRI > FDG >CSF)
HC-C = HC Converters; HC-NC = HC Non-Converters

Table 12: Classification result (AUC, sensitivity and specificity) for HC-C versus HC-NC of ADNI data with single modality and corresponding best multi-modality combination tested in DIAN.

Modality/Modalities	Mean AUC (95%CI)	Sensitivity, Specificity
MRI	64.65 (64.03, 65.27)	84.10, 52.05
FDG	62.79 (62.24, 63.35)	84.10, 34.46
PSY	68.17 (67.68, 68.67)	84.10, 61.00
CSF	49.57 (48.95, 50.19)	84.10, 10.67
P-C	58.22 (57.64, 58.79)	84.10, 22.49
F-P-C	63.28 (62.75, 63.82)	84.10, 35.13
M-F-P-C	62.69 (62.12, 63.26)	84.10, 36.06

3.3.3 Classification of MCI converters and non-converters in ADNI

We used all single modality models (MRI, FDG, PSY, CSF), the two-modality combination (PSY-CSF) and the three-modality combination (FDG-PSY-CSF) models with the best performance in DIAN, and all-modality combination model (MRI-FDG-PSY-CSF) to distinguish MCI converters and non-converters of ADNI dataset. The ROC curves and classification results (AUC, sensitivity, and specificity) of single- and multi- modality are shown in *Figure 23* and *Table 13*. Among the single modality models, the neuropsychological model had the best result (AUC = 83.24%), and CSF model (the best single model in DIAN) showed the AUC of 82.61% which was not better than the other single modalities. The two-modality combination PSY-CSF (best two-modality model in DIAN), achieved AUC of 88.38%. The three-modality combination FDG-PSY-CSF (best three-modality model established in DIAN), showed AUC of 89.12%. All-modality combination (MRI-FDG-PSY-CSF) achieved AUC of 88.66%.

Compared with each single modality model, FDG-CSF model had significantly better performance. The three-modality FDG-PSY-CSF showed significantly higher AUC than PSY-CSF model as well as all single-modality models. Four-modality-combination performed significantly better than all single-modality models but not better than two- and three-modality models.

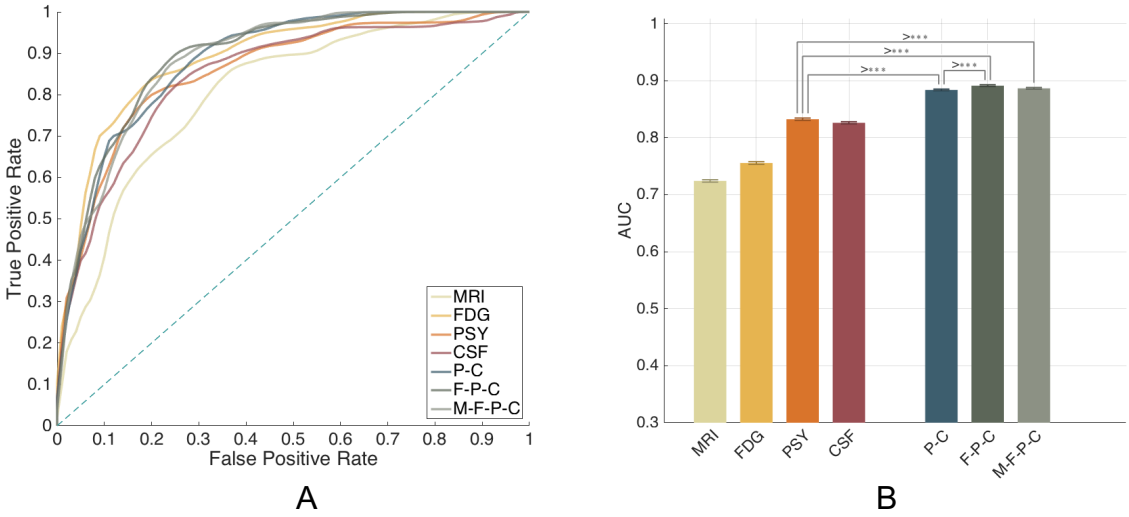

Figure 23. Classification result for MCI Converters (MCI-C) versus MCI non-Converters (MCI-NC) in SAD based on ADAD Naïve Bayes model with all four modalities: MR volume, FDG-PET, Psychological data and CSF biomarkers and the multi-modality combinations with best performance in ADAD. (a) ROC curves of classification: MCI-C vs. MC-NC. (b) AUC of classification: MCI-C vs. MCI-NC. The significant increasing of AUC is only marked for the highest one. (AUC: Psy >CSF > FDG-PET > MRI) MCI-C = MCI Converters; MCI-NC = MCI Non-Converters

Table 13: Classification result (AUC, sensitivity and specificity) for MCI-C versus MCI-NC of ADNI data with single modality and corresponding best multi-modality combination tested in DIAN.

Modality/Modalities	**Mean AUC (95%CI)**	**Sensitivity, Specificity**
MRI	72.41 (72.20, 72.63)	81.88, 53.76
FDG	75.56 (75.33, 75.79)	81.05, 54.83
PSY	83.24 (83.04, 83.44)	81.57, 69.92
CSF	82.61 (82.42, 82.80)	81.80, 76.06
P-C	88.38 (88.24, 88.52)	83.25, 76.99
F-P-C	89.12 (88.97, 89.28)	82.79, 82.42
M-F-P-C	88.66 (88.51, 88.80)	83.48, 80.92

3.3.4 Overview of results in ADNI

In general, the combinations of multiple modalities improved the classification result especially when the single modality performed relatively good (such as HC vs. AD, MCI converters vs. non-converters).

With the best models established in DIAN, classification results of ADNI were shown in *Figure 24* and *Table 14*. For the most difficult task, the AUC of distinguishing HC converter and non-converters was 62.69% using all four modalities. For predicting MCI conversion, the four-modality combination model showed AUC of 88.66%. To distinguish HC and AD, the best model in DIAN (FDG-CSF) achieved AUC of 94.80%.

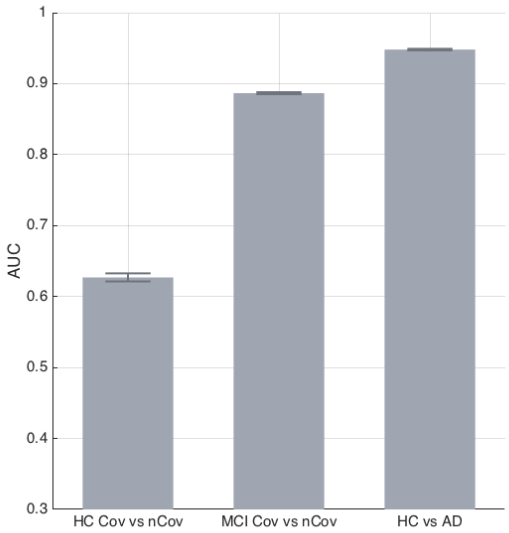

Figure 24. Classification result (AUC) for all three pairs of groups of ADNI data with corresponding best multi-modality combination tested in DIAN.

Table 14: Classification result (AUC, sensitivity and specificity) for all three pairs of groups of ADNI data with corresponding best multi-modality combination tested in DIAN.

Groups	Mean AUC (95%CI)	Sensitivity, Specificity	Best combinations in DIAN
HC Cov vs. nCov	62.69 (62.12, 63.26)	84.10, 36.06	MRI-FDG-PSY-CSF
MCI Cov vs. nCov	88.66 (88.51, 88.80)	83.48, 80.92	MRI-FDG-PSY-CSF
HC vs. AD	94.80 (94.69, 94.90)	89,11, 90.62	FDG-CSF

4 Discussion

In this study, we used Bayesian machine learning method: 1) to differentiate mutation and non-mutation carriers in ADAD and 2) to detect subjects at early risk-stages of SAD (HC to MCI/AD or MCI to AD converters)

For the classification of ADAD mutation and non-mutation carriers in the training sample (DIAN), the result showed that the classification accuracy increased with the progress of the disease (higher EYO). When combining multiple modalities, it achieved higher classification accuracy than a single modality. Moreover, classification with three-modality combination showed a better result than two-modality combination and combination of all the modalities showed the best result.

To predict the conversion in the early stage of SAD, we applied the same machine learning model established in DIAN (ADAD subjects). For predicting the conversion from MCI to AD, classification results achieved clinically relevant levels of accuracy when combining two or more modalities. Combinations of multiple modalities significantly improved the performance of the classifiers compared to single modality for distinguishing MCI converters and non-converters. For the more difficult task, predicting conversion from HC to MCI, the classification accuracy didn't achieve clinically relevant levels of accuracy.

For the discrimination between HC and AD, the same classification model establish in DIAN was applied. The AUC score ranged from 86% to 95% (sensitivity: 83% - 93%, specificity: 78% - 91%), supporting the plausibility of the established machine learning model.

4.1. Selected features

To improve the classification performance, we did feature selection for all four modalities (MRI, FDG-PET, CSF biomarkers and neuropsychological tests) using information gain (IG) method in DIAN across different EYO.

In MRI, feature selection was applied to cortical and subcortical volumes of 82 ROIs. Most informative features were selected to distinguish mutation and non-mutation carriers across the whole EYO range. The results showed volumes of paracentral lobule and postcentral gyrus were selected as most important features at more than 15 years before the onset of symptoms. Similar regions have been found showing a difference in amyloid deposition between mutation and non-mutation carriers at the same EYO and became more significant in the following years (Benzinger *et al.*, 2013). At the EYO $= (-15, -5]$, the same brain regions paracentral lobule and postcentral gyrus, with the addition of supramarginal gyrus were selected for classification. By EYO $= (-20, -10]$, selected features of grey matter volumes were left putamen and right lingual. Atrophy and ^{11}C-PiB accumulation in putamen have been reported in recent ADAD research with EYO = -10 and -15 respectively (Benzinger *et al.*, 2013; Gordon *et al.*, 2018). At the EYO interval $(-10, 0]$, accumbens and precuneus were selected as the most informative features. Consistent to our results, the same gray matter regions were found with greatest cortical thinning or volume loss at the same EYO interval (Benzinger *et al.*, 2013). From five years before the estimated year of onset, more features were selected including subcortical regions such as putamen, hippocampus, amygdala and cortical regions such as banks of the superior temporal sulcus, postcentral gyrus, cuneus and precuneus which were known to be vulnerable in AD. In ADAD, cortical thickness changes in precuneus, temporal lobe and subcortical regions when approaching to the year of symptom onset have been shown in the recent study (Gordon *et al.*, 2018), suggesting similar patterns between most informative features for

distinguishing mutation vs non-mutation carriers and brain structure changes in the progress of ADAD.

For FDG-PET, the most informative features were selected from SUVR scores of 82 ROIs across the entire EYO window. Selected features at more than 15 years before the onset of symptoms were shown in the caudal anterior cingulate cortex, fusiform and superior parietal cortex. Hypometabolism in parietal lobe has been observed in AD (Mosconi *et al.*, 2006). But the other regions selected in the very early stage have not been reported previously which may due to the subtle difference in this stage. By EYO $= (-20, -10]$, selected features of FDG-PET were hippocampus, amygdala and frontal pole. Hippocampus has been shown with reduction of glucose metabolism when approaching the symptom onset in ADAD (Benzinger *et al.*, 2013). Our result showed that hippocampus was also selected at the next EYO interval $(-15, -5]$ but not at the following EYO. Besides hippocampus, pallidum and accumbens of subcortical regions were selected at the EYO $= (-15, -5]$ which have been shown with a significant difference in PiB accumulation but not in FDG-PET between mutation carriers and non-carriers at almost the same time in ADAD (Benzinger *et al.*, 2013). At the EYO interval $(-10, 0]$, inferior parietal, cuneus, and isthmus-cingulate cortex were selected as the most informative features. Similar regions (cingulate cortex and parietal lobe) have been shown with a significant difference in hypometabolism between mutation carriers and noncarriers 10 years before symptom onset. More cortical features of FDG-PET were selected around the onset of symptoms including precuneus, isthmus-cingulate cortex, inferior and superior parietal cortex which were consistent with previous findings that these regions differ between carriers and non-carriers in ADAD and also showed metabolic reductions in SAD (Bateman *et al.*, 2012; Förster *et al.*, 2012; Gordon *et al.*, 2018).

The selected features from neuropsychological tests were relatively consistent across EYO. Boston Naming test was the most frequently selected feature among neuropsychological features which is a measure of naming ability and semantic memory and one of the most used

assessments of cognitive function in AD. Animal Category Fluency Test is a measure of semantic memory and executive function which was selected multiple times especially in the years before symptom onset. Fluency score has been used to predict MCI conversion and showed good performance in the recent study (Clark *et al.*, 2016). At the EYO interval $(-15, -5]$ and the years after symptom onset, Trail Making Tests (A&B) were selected for further classification. Trail Making Test is a measure of processing speed, visual–motor skills, and executive functions which has been shown with a relatively good result for predicting MCI conversion but not for differentiating HC and AD as a single predictor (Ewers *et al.*, 2012). As a measure of episodic memory, Word List Recall Test was selected at ten years before symptom onset and at the EYO interval $(-5, +5]$. However, in a recent study, it has been shown increasing difference between mutation and non-mutation carriers (CDR = 0) in the progress of ADAD and it has been reported with high accuracy for the classification of HC and AD (Ewers *et al.*, 2012; Storandt *et al.*, 2014). So the Word List Recall Test was not selected in the late stage of ADAD which remains unclear. MMSE, a widely used test to evaluate cognitive impairment and screen for possible dementia, was only selected at one EYO interval $(-10, 0]$. This may suggest that the global cognitive function did not show the significant difference between mutation and non-mutation carrier in most time during the disease progress which may due to late changes are usually shown in cognitive impairment. Among CSF biomarkers, phosphorylated Tau was selected as the informative feature across all EYO intervals, which may imply a significant difference between mutation and non-mutation carriers in the whole progress of ADAD. In recent research, concentrations of pTau have been shown significantly higher in mutation carriers than non-mutation carriers from 20 years before the estimated year of symptom onset which is consistent with our result (Fagan *et al.*, 2014b). As another selected feature of CSF biomarkers, Aβ was selected mainly in the later stage of ADAD compared to pTau which was from 15 years before symptom onset expect the EYO interval $(-5, +5]$. Similarly, $A\beta_{1\text{-}42}$ in mutation carriers was shown a

significant difference from those of non-mutation carriers from 10 years before symptom onset (Fagan *et al.*, 2014b). Even though $A\beta_{1\text{-}42}$ was selected across almost all time intervals from $\mathrm{EYO} = -15$, it still remains unclear why it was not selected at the EYO interval $(-5, +5]$ which also has been shown with a big difference between the two groups in the other study (Fagan *et al.*, 2014b). Longitudinal change of Tau in SAD and difference between the mutation and non-mutation carriers in ADAD was reported previously (Bateman *et al.*, 2011). However, in our study, total Tau was not selected across the whole range EYO.

4.2. Classification in DIAN

Previous studies have shown brain atrophy, glucose hypometabolism, abnormalities in CSF and cognitive impairment in the progress of ADAD. So we derived features from neuroimaging (MRI, FDG-PET), CSF biomarkers, and neuropsychological tests.

With the selected features, we trained the classification model for successive EYO intervals to predict the mutation status and the results showed increasing classification accuracy in the disease course.

For machine learning model establishing, we used Naïve Bayes algorithm using the features of MRI, FDG-PET, neuropsychological and CSF data. Naïve Bayes algorithm has been used in the prediction of AD previously and it has been proved to have good performance with limited size of training data and relatively low dimensional feature space which is fit for our training dataset of DIAN (Dyrba *et al.*, 2013; Dukart, Sambataro and Bertolino, 2015; Bhagya Shree and Sheshadri, 2018). In different stages of ADAD, classification model was trained with selected features of that EYO interval. As expected, classification accuracy increased with the EYO intervals approaching the symptom onset.

Among the classifications with single modality, CSF predictors yielded the highest, and in opposite neuropsychological predictors showed lowest AUC from 15 years before symptom

onset, which is in consistency with the temporal ordering of biomarkers in ADAD where the earliest change was CSF biomarkers and cognitive change was shown at the late stage of ADAD (Bateman *et al.*, 2012). At this EYO interval, AUC of MRI and FDG-PET measures were between neuropsychological test and CSF biomarkers. Previous studies reported that brain atrophy and hypometabolism could be detected following CSF biomarkers which could explain our result that CSF predictors performed better than MRI and FDG-PET in a very early stage of the disease. For the CSF predictors, classification result generally increased and better than the other single-modality models across almost all time intervals (except the EYO interval $(-15, -5]$). CSF predictors achieved a predictive accuracy from EYO interval $(-10, 0]$ which was earlier than FDG-PET, neuropsychological assessments and MRI. In general, the result of the single-modality classification models showed a trend toward increasing AUC when the time gets closer to the symptom onset, and it is consistent with the model of dynamic biomarkers in ADAD proposed in recent year (Bateman *et al.*, 2012). However, the measures used in our research were not exactly same as those of the temporal model of ADAD (e.g. gray matter volumes of multiple regions used in our study and only hippocampus volume used in the temporal model), so the minor difference still remains.

For the two-modality predictors, there were 6 different combinations of MRI, FDG-PET, neuropsychological tests and CSF biomarkers. Comparing with single-modality predictors, we found that the best two-modality combinations were not always the combination of the best single modality which was also shown in the previous study (Dukart, Sambataro and Bertolino, 2015). Combining MRI and CSF biomarkers, the AUC became clinical relevant from the EYO interval of $(-20, -10]$ which was earlier than every single modality predictor. In general, the classification result of either the best combination or averaging all six combinations were better than those of single modality which was shown in *Table 9* and *Figure 26*.

For the three-modality predictors, combining MRI, FDG-PET and CSF biomarkers achieved the highest AUC at most EYO intervals. Generally, the average of all three-modality combinations was better than those of 2-modality combinations which were shown in *Figure 26*. The best combination models with three modalities were better than two-modality combination for most EYO intervals. When the predictors with every single modality showed relatively high AUC especially in the late stage of ADAD, the multi-modality model may gain extra benefits through combination. In comparison with the best model of 3-modality combinations at each EYO interval, 4-modality combination model performed better only at the EYO interval of $(-5, +5]$. However, the AUC of the model combining all modalities was higher than the average of all three-modality combinations across all EYO intervals. In general, combining multiple modalities showed a gain over the model of fewer modalities for the discrimination between mutation and non-mutation carriers in ADAD (*Figure 26*). Mutation carriers of ADAD can be tracked decades before the symptom onset. Importantly, clear AD-related brain changes can be detected in the early stage of ADAD. Multi-modality modals to predict conversion of AD have been used in previous studies (Ewers *et al.*, 2012; Segovia *et al.*, 2014; Dukart, Sambataro and Bertolino, 2015; Korolev, Symonds and Bozoki, 2016). All of these studies established the model based on SAD related dataset which may include non-AD-specific pathology for the subjects who are in the early stage of AD. Hence, we used the machine learning model based on pure AD-related features extracted from DIAN participants to predict the conversion of SAD. Before further applying the prediction model to ADNI dataset, we evaluated the models of different modality predictors at different EYO intervals in DIAN. Considering the model need to be relativity accurate and robust, we thus chose the prediction model trained for EYO interval of $(-5, \mathrm{Inf})$ which is above the clinically relevant accuracy and similar to the early stage of SAD. The EYO interval of $(-5, \mathrm{Inf})$ is combined of two intervals to enlarge the sample size and gain more power for the model.

For MRI, the selected features included hippocampus, putamen, amygdala, precuneus, postcentral gyrus, and cuneus. Difference between mutation and non-mutation carriers in the similar regions and timepoint in the course of ADAD was shown in the recent study where the significant difference was observed in postcentral gyrus , precuneus, cuneus and hippocampus (Gordon *et al.*, 2018). Consistent brain regions (inferior and superior parietal cortex, precuneus) were also found in FDG-PET both in our and Gordon's study. For neuropsychological tests, the most discriminative features were Boston Naming Test, Trail Making Test (A&B) and Word List Immediate Recall. For CSF biomarkers, Aβ and pTau were selected as the most informative features. Our classification results showed the AUCs of all single and multiple modalities predictors were above the clinically relevant accuracy for distinguishing mutation and non-mutation carriers at the EYO interval of $(-5, +Inf)$ which met one of the prerequisites: an accurate predictive model. Based the classification results validated within DIAN dataset, the best models among different combinations were selected for further validation in ADNI dataset.

4.3. Classification in ADNI

Applied with the best models of the different number of modalities established in DIAN, classification in the more common SAD (ADNI dataset) showed high accuracy for distinguishing HC vs. AD and MCI converters vs. non-converters.

For the classification of HC and AD subjects, the CSF model showed the best result among the single-modality models. When FDG-PET was added, the combination of CSF and FDG-PET yielded significant a better result than the single-modality model. Compared to two-modality model, combining MRI, FDG-PET and CSF biomarker achieved significant higher AUC of 96.00% (sensitivity = 93.35%, specificity = 91.40%). Similarly, a combination of these three modalities was reported in a previous study with an accuracy of 92.0% which was

significantly better than the single modality. It should be noted that the psychological data was not included for the classification of HC and AD, since it provided diagnostic information which may lead to circularity.

For the prediction of conversion in the early stage of AD, the classification for both HC and MCI converters versus non-converters performed best when using neuropsychological data among the single predictors. This result is consistent with the previous relevant study which reported higher AUC using cognitive measures than MRI and CSF data to predict HC to MCI at 5 years (Albert et al., 2018). With multi-modalities, the classification for MCI converters versus non-converters in SAD showed the best result when using FDG-PET, neuropsychological and CSF data which increased significantly compared to single and two-modality models with AUC of 89.16% which is above the clinical relevant accuracy. This result was comparable with recent studies which reported classification accuracy between 80% to 90% for discrimination of MCI converters and non-converters combining three modalities (Segovia *et al.*, 2014; Cheng *et al.*, 2015; Dukart, Sambataro and Bertolino, 2015; Moradi *et al.*, 2015; Suk, Lee and Shen, 2015; Wang *et al.*, 2016; Tong *et al.*, 2017). However, different modality combinations were used in different studies which limited the comparability of accuracies.

Combination of all four modalities didn't show a better result than three modalities, indicating the model didn't gain more useful information for the prediction of MCI conversion by adding MRI features. Combination of different modalities didn't benefit for HC converters versus non-converters where none of the multi-modalities performed better than the single modality of neuropsychological data. There were very few studies in recent years predicted the conversion of healthy subjects which may due to limited of follow-up for HC and thus the sample size (the number of subjects who developed to MCI/AD) was not sufficient for prediction. In our study, there was only 14 HC subjects who converted within 36 months and the limited the sample size may lead to lacking robustness. There was one recent study

reported relatively high AUC for predicting progression from HC to MCI at 5 years based on the Cox proportional hazards model. However, with machine learning algorithms, there was no study showing the prediction of HC conversion so far (Pellegrini *et al.*, 2018). Though many previous studies used multi-modality machine learning models for the prediction in the early stage of AD in previous studies, most studies tested the prediction model using the cross-validation method. This means training and testing data were not independent and it may increase the risk of overfitting. Our current study used fully independent training (DIAN dataset) and testing set (ADNI dataset). Even though ADAD only accounts for a small proportion of AD, because of similar pathological and clinical features, it has a critical role to better understand the pathophysiology in the preclinical stage of SAD.

5 Conclusion

With the use of neuroimaging, psychological and CSF biomarkers, we did classification for successive EYO intervals to predict the mutation status and the results showed increasing classification accuracy in the disease course for both single modality and multi-modality models. Considering the model need to be relativity accurate and robust, we thus chose the prediction model trained for EYO interval of $(-5, \mathrm{Inf})$ which is above the clinically relevant accuracy and similar to the early stage of SAD.

Applied with the best models established in DIAN, classification in the more common SAD (ADNI dataset) showed high accuracy for distinguishing HC vs. AD and MCI converters vs. non-converters. Combining multiple modalities, the prediction model yielded a significant better result than the single-modality model. Particularly, the classification for MCI converters versus non-converters in SAD showed the best result when using FDG-PET, neuropsychological and CSF data which increased significantly compared to single and two-modality models with AUC of 89.12% which is above the clinical relevant accuracy. In summary, the machine learning model established in DIAN achieved predictive accuracy to predict the conversion in the early stage of SAD. Moreover, the current study provides good interpretability to understand SAD through autosomal-dominant AD.

6 Supplement

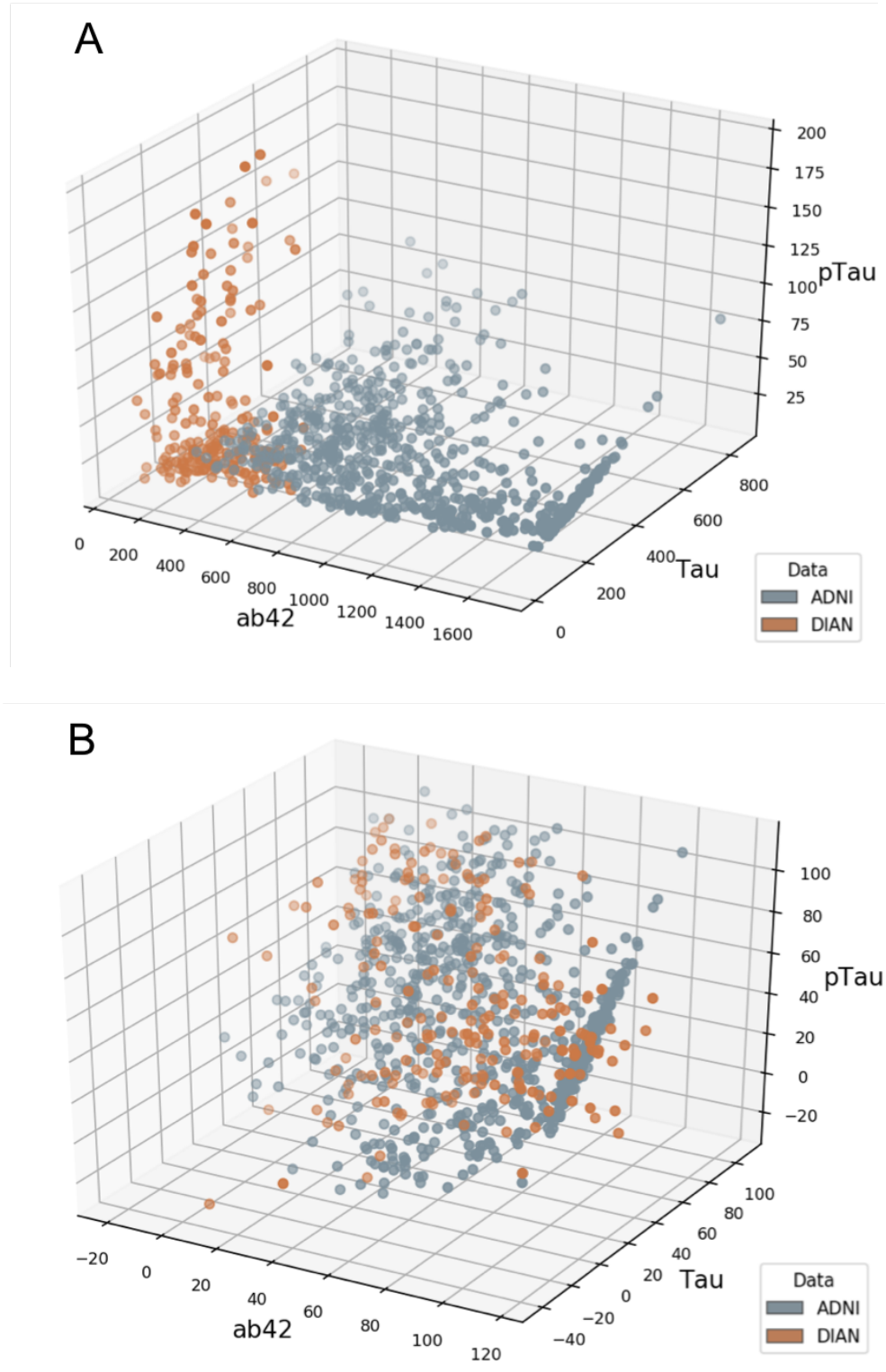

Figure 25 (A) 3D scatter plot of CSF data for the group comparison (ADNI vs DIAN) (A) before and (B) after normalization and standardization.

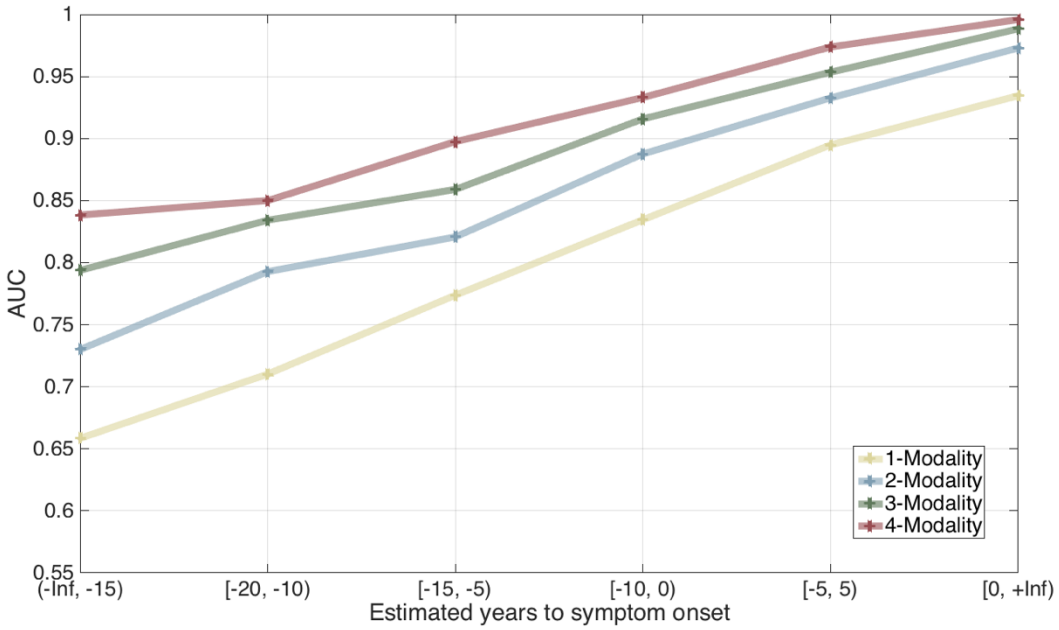

Figure 26 Mean AUC for the discrimination of mutation carriers vs. non-mutation carriers of single modality, two-modality combination, three-modality and all combination across different EYO intervals.

7 References

Bateman, R. J. *et al.* (2011) 'Autosomal-dominant Alzheimer ' s disease : a review and proposal for the prevention of Alzheimer ' s disease', pp. 1–13.

Bateman, R. J. *et al.* (2012) 'Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease', *New England Journal of Medicine*, 367(9), pp. 795–804. doi: 10.1056/NEJMoa1202753.

Beheshti, I., Demirel, H. and Matsuda, H. (2017) 'Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm', *Computers in Biology and Medicine*. Elsevier Ltd, 83(February), pp. 109–119. doi: 10.1016/j.compbiomed.2017.02.011.

Benzinger, T. L. S. *et al.* (2013) 'Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease', *Proceedings of the National Academy of Sciences*, 110(47), pp. E4502–E4509. doi: 10.1073/pnas.1317918110.

Bhagya Shree, S. R. and Sheshadri, H. S. (2018) 'Diagnosis of Alzheimer's disease using Naive Bayesian Classifier', *Neural Computing and Applications*. Springer London, 29(1), pp. 123–132. doi: 10.1007/s00521-016-2416-3.

Bicego, M. and Baldo, S. (2016) 'Properties of the Box–Cox transformation for pattern classification', *Neurocomputing*. Elsevier, 218, pp. 390–400.

Blennow, K. *et al.* (2010) 'Cerebrospinal fluid and plasma biomarkers in Alzheimer disease', *Nature Reviews Neurology*. Nature Publishing Group, 6(3), p. 131.

Blennow, K., de Leon, M. J. and Zetterberg, H. (2006) 'Alzheimer's disease', *Lancet*, 368(9533), pp. 387–403. doi: 10.1016/S0140-6736(06)69113-7.

Braak, H. *et al.* (2006) 'Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry', *Acta neuropathologica*. Springer, 112(4), pp. 389–404.

Braak, H. and Braak, E. (1991) 'Neuropathological stageing of Alzheimer-related changes', *Acta neuropathologica*. Springer, 82(4), pp. 239–259.

Buerger, K. *et al.* (2006) 'CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease', *Brain*, 129(11), pp. 3035–3041. doi: 10.1093/brain/awl269.

Cabral, C. *et al.* (2015) 'Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages', *Computers in Biology and Medicine*, 58, pp. 101–109. doi: 10.1016/j.compbiomed.2015.01.003.

Campion, D. *et al.* (1999) 'Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum.', *American journal of human genetics*, 65(3), pp. 664–70. doi: 10.1086/302553.

Chen, X., Wei, K. and Liu, M. (2015) 'Group Sparse Representation for Prediction of MCI Conversion to AD', in *International Conference on Intelligent Computing*. Springer, pp. 510–519.

Cheng, B. *et al.* (2015) 'Domain Transfer Learning for MCI Conversion Prediction', *IEEE Transactions on Biomedical Engineering*, 62(7), pp. 1805–1817. doi: 10.1109/TBME.2015.2404809.

Choi, H. and Jin, K. H. (2018) 'Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging', *Behavioural Brain Research*, 344(2010), pp. 103–109. doi: 10.1016/j.bbr.2018.02.017.

Christina, P. (2018) 'World Alzheimer's report 2018', *Alzheimer's disease Internations: world Alzheimer report 2018*, pp. 1–48. doi: 10.1111/j.0033-0124.1950.24_14.x.

Clark, D. G. *et al.* (2016) 'Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment', *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. Elsevier Inc., 2, pp. 113–122. doi: 10.1016/j.dadm.2016.02.001.

Cruts, M. and Van Broeckhoven, C. (1998) 'Molecular Genetics of Alzheimer's Disease', *Annals of Medicine*, 30, pp. 560–565.

Desikan, R. S. *et al.* (2006) 'An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest', *NeuroImage*, 31(3), pp. 968–980. doi: 10.1016/j.neuroimage.2006.01.021.

Dickerson, B. C. *et al.* (2001) 'MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease☆', *Neurobiology of aging*. Elsevier, 22(5), pp. 747–754.

Dickerson, B. C. and Wolk, D. (2013) 'Biomarker-based prediction of progression in MCI: comparison of AD-signature and hippocampal volume with spinal fluid amyloid-β and tau', *Frontiers in aging neuroscience*. Frontiers, 5, p. 55.

Dukart, J., Sambataro, F. and Bertolino, A. (2015) 'Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers', *Journal of Alzheimer's Disease*, 49(4), pp. 1143–1159. doi: 10.3233/JAD-150570.

Dyrba, M. *et al.* (2013) 'Robust Automated Detection of Microstructural White Matter Degeneration in Alzheimer's Disease Using Machine Learning Classification of Multicenter DTI Data', *PLoS ONE*, 8(5). doi: 10.1371/journal.pone.0064925.

Dyrba, M. *et al.* (2015) 'Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data', *Journal of Neuroimaging*, 25(5), pp. 738–747. doi: 10.1111/jon.12214.

Eisenstein, E. L. and Alemi, F. (1993) 'An evaluation of factors influencing Bayesian learning systems.', *Proceedings. Symposium on Computer Applications in Medical Care*, pp. 485–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8130520.

Ewers, M. *et al.* (2011) 'Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia', *Trends in Neurosciences*. doi: 10.1016/j.tins.2011.05.005.

Ewers, M. *et al.* (2012) 'Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance', *Neurobiology of Aging*. Elsevier Inc., 33(7), p. 1203–1214.e2. doi: 10.1016/j.neurobiolaging.2010.10.019.

Fagan, A. M. *et al.* (2014a) 'Longitudinal Change in CSF Biomarkers in Autosomal-Dominant Alzheimer's Disease', *Science Translational Medicine*, 6(226), p. 226ra30-226ra30. doi: 10.1126/scitranslmed.3007901.

Fagan, A. M. *et al.* (2014b) 'Longitudinal Change in CSF Biomarkers in Autosomal-Dominant Alzheimer's Disease', *Science Translational Medicine*, 6(226), p. 226ra30-226ra30. doi: 10.1126/scitranslmed.3007901.

Fischl, B. *et al.* (2002) 'Whole Brain Segmentation : Neurotechnique Automated Labeling of Neuroanatomical Structures in the Human Brain', 33, pp. 341–355.

Fischl, B. (2012) 'FreeSurfer', *NeuroImage*, 62, pp. 774–781. doi: 10.1016/j.neuroimage.2012.01.021.

Fodero-Tavoletti, M. T. *et al.* (2011) '18F-THK523: A novel in vivo tau imaging ligand for Alzheimer's disease', *Brain*, 134(4), pp. 1089–1100. doi: 10.1093/brain/awr038.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975) '"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.', *Journal of psychiatric research*, 12(3), pp. 189–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1202204.

Förster, S. *et al.* (2012) 'Regional expansion of hypometabolism in Alzheimer's disease follows amyloid deposition with temporal delay', *Biological psychiatry*. Elsevier, 71(9), pp. 792–797.

Frisoni, G. B. *et al.* (2008) 'Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla', *Brain*. Oxford University Press, 131(12), pp. 3266–3276.

Frölich, L. *et al.* (2017) 'Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer's dementia', *Alzheimer's Research and Therapy*. Alzheimer's Research & Therapy, 9(1), pp. 1–15. doi: 10.1186/s13195-017-0301-7.

Furst, A. J. *et al.* (2012) 'Cognition, glucose metabolism and amyloid burden in Alzheimer's disease', *Neurobiology of aging*. Elsevier, 33(2), pp. 215–225.

Goodglass, H. and Kaplan, E. (1983) 'Boston Diagnostic Aphasia Examination', in. Lea and Febiger, Philadelphia.

Gordon, B. A. *et al.* (2018) 'Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study', *The Lancet Neurology*, 17(3), pp. 211–212. doi: 10.1016/S1474-4422(18)30028-0.

Guerrero, R. *et al.* (2014) 'Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO', *NeuroImage*. Elsevier Inc., 94, pp. 275–286. doi: 10.1016/j.neuroimage.2014.03.036.

Hall, M. A. and Holmes, G. (2003) 'Benchmarking attribute selection techniques for discrete class data mining', *IEEE Transactions on Knowledge and Data engineering*. IEEE, 15(6), pp. 1437–1447.

Hor, S. and Moradi, M. (2016) 'Learning in data-limited multimodal scenarios: Scandent decision forests and tree-based features', *Medical Image Analysis*. Elsevier B.V., 34, pp. 30–41. doi: 10.1016/j.media.2016.07.012.

Hu, K. *et al.* (2015) 'Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis', *Neurocomputing*. Elsevier, 175(PartA), pp. 132–145. doi: 10.1016/j.neucom.2015.10.043.

Iqbal, K. *et al.* (2005) 'Tau pathology in Alzheimer disease and other tauopathies', *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*. Elsevier, 1739(2–3), pp. 198–210.

Irani, K. and Fayyad, U. (1993) 'Multi-lnterval Discretization of Continuous-Valued Attributes for Classification learning', *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1022–1027. doi: 10.1109/TKDE.2011.181.

Jack, C. R. *et al.* (2010) 'Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade', *The Lancet Neurology*, 9(1), pp. 119–128. doi: 10.1016/S1474-4422(09)70299-6.

Jack, C. R. *et al.* (2010) 'Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 6(3), pp. 212–20. doi: 10.1016/j.jalz.2010.03.004.

Jack, C. R. *et al.* (2013) 'Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers', *The Lancet Neurology*, 12(2), pp. 207–216. doi: 10.1016/S1474-4422(12)70291-0.

Jack, C. R. *et al.* (2015) 'Age, sex, and APOE ϵ4 effects on memory, brain structure, and β-Amyloid across the adult life Span', *JAMA Neurology*, 72(5), pp. 511–519. doi: 10.1001/jamaneurol.2014.4821.

Jack, R. *et al.* (1989) 'Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults.', *Radiology*, pp. 549–554.

Jansen, W. J. *et al.* (2015) 'Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis', *Jama*. American Medical Association, 313(19), pp. 1924–1938.

Jie, B. *et al.* (2015) 'Manifold regularized multitask feature learning for multimodality disease classification', *Human Brain Mapping*, 36(2), pp. 489–507. doi: 10.1002/hbm.22642.

Khazaee, A., Ebrahimzadeh, A. and Babajani-Feremi, A. (2016) 'Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease, *Brain Imaging and Behavior*, 10(3), pp. 799–817. doi: 10.1007/s11682-015-9448-7.

Klunk, W. E. *et al.* (2004) 'Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B', *Annals of Neurology*, 55(3), pp. 306–319. doi: 10.1002/ana.20009.

Klunk, W. E. *et al.* (2015) 'The Centiloid project: Standardizing quantitative amyloid plaque estimation by PET', *Alzheimer's and Dementia*. Elsevier Inc, 11(1), p. 1–15.e4. doi: 10.1016/j.jalz.2014.07.003.

Komlagan, M. *et al.* (2014) 'Anatomically constrained weak classifier fusion for early detection of Alzheimer's disease', in *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 141–148.

Korolev, I. O., Symonds, L. L. and Bozoki, A. C. (2016) 'Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification', *PLoS ONE*, 11(2), pp. 1–25. doi: 10.1371/journal.pone.0138866.

Lee, H. *et al.* (2005) 'Tau phosphorylation in Alzheimer's disease: pathogen or protector?', *Trends in molecular medicine*. Elsevier, 11(4), pp. 164–169.

Lin, W. *et al.* (2018) 'Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment', *Frontiers in Neuroscience*, 12(November), pp. 1–13. doi: 10.3389/fnins.2018.00777.

Liu, F. *et al.* (2014) 'Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification', *NeuroImage*. Elsevier Inc., 84, pp. 466–475. doi: 10.1016/j.neuroimage.2013.09.015.

Liu, M. *et al.* (2016) 'Inherent Structure-Based Multiview Learning with Multitemplate Feature Representation for Alzheimer's Disease Diagnosis', *IEEE Transactions on Biomedical Engineering*, 63(7), pp. 1473–1482. doi: 10.1109/TBME.2015.2496233.

Long, X. *et al.* (2017) 'Prediction and classification of Alzheimer disease based on quantification of MRI deformation', *PLoS ONE*, 12(3), pp. 1–19. doi: 10.1371/journal.pone.0173372.

Mattson, M. P. (2004) 'Pathways towards and away from Alzheimer's disease', *Nature*, 431(7004), pp. 107–107. doi: 10.1038/nature02940.

Mayeux, R. *et al.* (2003) 'Plasma Aβ40 and Aβ42 and Alzheimer's disease: relation to age, mortality, and risk', *Neurology*. AAN Enterprises, 61(9), pp. 1185–1190.

Mayeux, R. *et al.* (2011) 'Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment - Part 1', *Alzheimer's and Dementia*. Elsevier Ltd, 7(1), pp. 15–34. doi: 10.1016/j.jalz.2010.11.005.

Minhas, S. *et al.* (2017) 'A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data', *IEEE Journal of Biomedical and Health Informatics*, 21(5), pp. 1403–1410. doi: 10.1109/JBHI.2016.2608998.

Misra, C., Fan, Y. and Davatzikos, C. (2009) 'Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI', *Neuroimage*. Elsevier, 44(4), pp. 1415–1422.

Moradi, E. *et al.* (2015) 'Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects', *NeuroImage*. Elsevier Inc., 104, pp. 398–412. doi: 10.1016/j.neuroimage.2014.10.002.

Morris, J. C. *et al.* (1989) 'The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease.', *Neurology*, 39(9), pp. 1159–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2771064.

Mosconi, L. *et al.* (2006) 'Hypometabolism exceeds atrophy in presymptomatic early-onset familial Alzheimer's disease', *Journal of Nuclear Medicine*. Society of Nuclear Medicine, 47(11), pp. 1778–1786.

Moulder, K. L. *et al.* (2013) 'Dominantly Inherited Alzheimer Network: facilitating research and clinical trials', *Alzheimer's research & therapy*. BioMed Central, 5(5), p. 48.

Niu, H. *et al.* (2017) 'Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis', *Neurología (English Edition)*. Sociedad Española de Neurología, 32(8), pp. 523–532. doi: 10.1016/j.nrleng.2016.02.009.

van Oijen, M. *et al.* (2006) 'Plasma Aβ1–40 and Aβ1–42 and the risk of dementia: a prospective case-cohort study', *The Lancet Neurology*. Elsevier, 5(8), pp. 655–660.

Ojala, M. and Garriga, G. C. (2009) 'Permutation Tests for Studying Classifier Performance', in *2009 Ninth IEEE International Conference on Data Mining*. IEEE, pp. 908–913. doi: 10.1109/ICDM.2009.108.

Pellegrini, E. *et al.* (2018) 'Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review', *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, pp. 519–535. doi: 10.1016/j.dadm.2018.07.004.

Plant, C. *et al.* (2010) 'Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease', *NeuroImage*. Elsevier Inc., 50(1), pp. 162–174. doi: 10.1016/j.neuroimage.2009.11.046.

Prince, M. *et al.* (2015) 'World Alzheimer Report 2015: The Global Impact of Dementia | Alzheimer's Disease International'.

Prince, M. *et al.* (2016) 'World Alzheimer Report 2016 Improving healthcare for people

living with dementia. Coverage, Quality and costs now and in the future', *Alzheimer's Disease International (ADI)*, pp. 1–140. doi: 10.13140/RG.2.2.22580.04483.

Quinlan, J. R. (2014) *C4. 5: programs for machine learning*. Elsevier.

Reitan, R. M. and Wolfson, D. (1985) *The Halstead-Reitan Neuropsychological Test Battery*. Neuropsychological Publishing Group, Tucson.

Ringman, J. M. *et al.* (2012) 'Cerebrospinal fluid biomarkers and proximity to diagnosis in preclinical familial Alzheimer's disease', *Dementia and geriatric cognitive disorders*. Karger Publishers, 33(1), pp. 1–5.

Rosen, W. G., Mohs, R. C. and Davis, K. L. (1984) 'A new rating scale for Alzheimer's disease.', *The American journal of psychiatry*, 141(11), pp. 1356–64. doi: 10.1176/ajp.141.11.1356.

Roussel, O. G., Ma, Y. and Evans, A. C. (1998) 'Correction for Partial Volume Effects in PET : Principle and Validation', 39(5).

Ryan, N. S. and Rossor, M. N. (2010) 'Correlating familial Alzheimer's disease gene mutations with clinical phenotype', *Biomarkers in medicine*. Future Medicine, 4(1), pp. 99–112.

Segovia, F. *et al.* (2014) 'Combining PET Images and neuropsychological test data for automatic diagnosis of alzheimer's disease', *PLoS ONE*, 9(2), pp. 1–8. doi: 10.1371/journal.pone.0088687.

Shattuck, D. W. *et al.* (2008) 'Construction of a 3D probabilistic atlas of human cortical structures', *Neuroimage*. Elsevier, 39(3), pp. 1064–1080.

Shaw, L. M. *et al.* (2009) 'Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects', *Annals of Neurology*, 65(4), pp. 403–413. doi: 10.1002/ana.21610.

Silverman, D. H. S. *et al.* (2001) 'Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome', *Jama*. American Medical Association, 286(17), pp. 2120–2127.

Sørensen, L. *et al.* (2016) 'Early detection of Alzheimer's disease using MRI hippocampal texture', *Human Brain Mapping*, 37(3), pp. 1148–1161. doi: 10.1002/hbm.23091.

Sperling, R. A. *et al.* (2011) 'Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease', *Alzheimer's and Dementia*. Elsevier Ltd, 7(3), pp. 280–292. doi: 10.1016/j.jalz.2011.03.003.

Storandt, M. *et al.* (2014) 'Clinical and psychological characteristics of the initial cohort of the Dominantly Inherited Alzheimer Network (DIAN).', *Neuropsychology*, 28(1), pp. 19–29. doi: 10.1037/neu0000030.

Suk, H.-I., Lee, S.-W. and Shen, D. (2014) 'Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis', *NeuroImage*, 101, pp. 569–582. doi: 10.1016/j.neuroimage.2014.06.077.

Suk, H.-I., Lee, S.-W. and Shen, D. (2015) 'Latent feature representation with stacked auto-encoder for AD/MCI diagnosis', *Brain Structure and Function*, 220(2), pp. 841–859. doi: 10.1007/s00429-013-0687-3.

Sunderland, T. *et al.* (2003) 'Decreased β-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease', *Jama*. American Medical Association, 289(16), pp. 2094–2103.

Tapiola, T. *et al.* (2009) 'Cerebrospinal fluid β-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain', *Archives of neurology*. American Medical Association, 66(3), pp. 382–389.

Terry, R. D. *et al.* (1991) 'Physical Basis of Cognitive Alterations in Alzheimer ' s Disease : Synapse h s s Is the Major Correlate of Cognitive Impairment', *Annals of Neurology*, 30(4), pp. 572–580.

Tong, T. *et al.* (2017) 'A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease', *IEEE Transactions on Biomedical Engineering*, 64(1), pp. 155–165. doi: 10.1109/TBME.2016.2549363.

Touchon, J. and Ritchie, K. (1999) 'Prodomal Cognitive Disorder in Alzheimer Disease', 563(December 1998), pp. 556–563.

Tzourio-Mazoyer, N. *et al.* (2002) 'Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain', *Neuroimage*. Elsevier, 15(1), pp. 273–289.

Villemagne, V. L. *et al.* (2013) 'Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study', *The Lancet Neurology*. Elsevier Ltd, 12(4), pp. 357–367. doi: 10.1016/S1474-4422(13)70044-9.

Vogt, N. (2018) 'Machine learning in neuroscience', *Nature Methods*. Nature Publishing Group, 15(1), p. 33.

Vu, M.-A. T. *et al.* (2018) 'A shared vision for machine learning in neuroscience', *Journal of Neuroscience*. Soc Neuroscience, pp. 508–517.

Walsh, D. M. and Selkoe, D. J. (2007) 'Aβ oligomers - A decade of discovery', *Journal of Neurochemistry*, 101(5), pp. 1172–1184. doi: 10.1111/j.1471-4159.2006.04426.x.

Wang, P. *et al.* (2016) 'Multimodal classification of mild cognitive impairment based on partial least squares', *Journal of Alzheimer's Disease*, 54(1), pp. 359–371. doi: 10.3233/JAD-160102.

Weintraub, S. *et al.* (2009) 'The Alzheimer's Disease Centers' Uniform Data Set (UDS): The Neuropsychological Test Battery', 23(2), pp. 91–101. doi: 10.1097/WAD.0b013e318191c7dd.The.

Wilson, R. *et al.* (2012) 'The natural history of cognitive decline in Alzheimer's disease', *Psychology and Aging*, 27(4), pp. 1008–17.

Xu, L. *et al.* (2015) 'Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment', *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 122(2), pp. 182–190. doi: 10.1016/j.cmpb.2015.08.004.

Zhu, X., Suk, H. Il and Shen, D. (2014) 'A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis', *NeuroImage*. Elsevier Inc., 100, pp. 91–105. doi: 10.1016/j.neuroimage.2014.05.078.

8 Acknowledgements

I would like to thank all people who helped me during my PhD study.

First and most importantly, I want to thank my supervisor Prof. Dr. Michael Ewers for the support and guidance during the past years. I have been grateful for the opportunities to carry out the interdisciplinary projects which I really learned a lot. I appreciated my supervisor's door was always open for questions and discussions and provided me helpful and inspiring advices.

I would like to thank to my colleague Miguel Ángel Araque Caballero who helped me with parts of the data preprocessing and give me some valuable advices during my PhD study. And also thank to Nicolai Franzmeier and Julia Neitzel for their helps with my projects and dissertation. Many thanks also to my current and previous colleagues Ruth Adam, Benno Gesierich, Lee Simon-Vermot, Yifei Zhang, Alexander Taylor, Ebru Baykara, Marek Konieczny, and everyone in the Lab for Brain Imaging and Biomarker and the Lab for Vascular Cognitive Impairment in ISD for their help and suggestions during my PhD study.

I would also like to thanks to my collaboration partner Wei Ye from the Institute for Informatics of LMU for providing a lot of suggestions of methods and the important technical support of the project.

Thanks to my parents for their support and understanding during these years. There won't be any difficulties which cannot be overcomed with their encouragements. And I also want to thank to my husband, for his acccompy and support in these years.

Finally, I want to say thanks again to everyone who helped no matter for my PhD study or my daily life. I'm deeply grateful for the help!